THE PHYLOGEOGRAPHY AND POPULATION GENETIC STRUCTURE OF LEAST CISCO (*COREGONUS SARDINELLA*) IN ALASKA

By

Veronica M. Padula

RECOMMENDED:

Dr. Anthony J. Gharrett

Dr. Douglas Causey, Co-Advisory Committee Chair

Dr. J. Andres Lopez, Co-Advisory Committee Chair

Dr. Shannon Atkinson, Chair, Graduate Program, Fisheries Division

APPROVED:

Dr. Michael Castellini, Dean, School of Fisheries and Ocean Sciences

Dr. John Eichelberger, Dean of the Graduate School

1/13/14

Date

THE PHYLOGEOGRAPHY AND POPULATION GENETIC STRUCTURE OF LEAST CISCO (*COREGONUS SARDINELLA*) IN ALASKA

A

THESIS

Presented to the Faculty

of the University of Alaska Fairbanks

in Partial Fulfillment of the Requirements

for the Degree of

MASTER OF SCIENCE

By

Veronica M. Padula, B.A.

Fairbanks, Alaska

December 2013

Abstract

The least cisco (*Coregonus sardinella*) is a whitefish species broadly distributed across the Arctic regions of Russia, Alaska, and Canada, and little is known about the genetic relationships among groups within this species. We investigated the genetic relationships among least cisco on two landscape scales. On a broader landscape scale, we investigated the relationships among populations across the state of Alaska by comparing mitochondrial DNA (mtDNA) sequences. On a finer landscape scale, we investigated the relationships among least cisco populations in closely located lakes on the Arctic Coastal Plain by comparing microsatellite DNA haplotypes. Data from mtDNA suggest that least cisco are relatively diverse across Alaska, with 68 unique haplotypes found in 305 individuals and a large proportion of genetic variation is shared across Alaska, but this variation is not homogeneously distributed across all regions and for all haplotype groups. Interpretation of microsatellite data was limited. Overall, the data suggest that least cisco populations are currently isolated from one another. Isolation also occurred historically, accounting for divergence among major clades. But general recontact events occurred as isolated populations migrated and colonized new habitats, accounting for the heterogeneity found across Alaska. Ultimately, Alaskan least cisco may have functioned as a metapopulation historically, but present populations are too isolated to be considered a metapopulation today.

Table of Contents

List of Figures

List of Tables

General introduction

This thesis is organized into three main chapters. The first presents a review of the literature, concepts and methods used to develop and complete the studies presented in the subsequent two chapters. The review consists of three major sections: genetics and geography, least cisco (*Coregonus sardinella*) biology, and possible effects of climate change on least cisco. The genetics and geography section reviews the concepts used in the study of phylogeography and population genetics. The least cisco section is a summary of background information on the evolution of salmonids (salmons, chars, trouts, whitefishes and their relatives) and coregonins (the whitefish and cisco group of salmonids), and on the life history and distribution of least cisco in Alaska. That section concludes by addressing the importance of whitefish to Native Alaskan subsistence culture. The final section of the first chapter summarizes current views of the potential impacts of climate change on the hydrography of Alaska and includes suggestions for possible roles for freshwater fish as indicators of change.

Chapter 2 presents a study of the phylogeography of least cisco in Alaska. This study is based on an examination of the spatial distribution of mitochondrial genetic diversity across the state. Results from this study showed that least cisco are genetically diverse across Alaska, with 68 unique mitochondrial haplotypes found among 305 individuals sampled. There is evidence of incipient genetic differentiation between the major regions included in the study. Overall, the data show that a large proportion of genetic variation is shared across Alaska, but this variation is not homogenously distributed across all regions and for all haplotype groups.

Chapter 3 consists of a population genetics study focusing on least cisco distributed in freshwater lakes across the Arctic Coastal Plain of Alaska. Little is known about these populations, including life history patterns and population genetic structure. To better understand the genetic structure of these populations, I examined microsatellite genotypes from eight Arctic Coastal Plain lakes. Conclusions from this study are constrained by considerations of data quality that are discussed in the chapter.

Overall, the results from analyses presented in this thesis show substantial level of genetic diversity in least cisco in Alaska and significant levels of genetic differentiation between populations in different regions of the state.

Chapter 1: Literature review

Introduction

Here, I review the concepts and methods used in this thesis, including phylogeography, population genetics and the biology of least cisco. Studies of phylogeography and population genetics both provide information about the history of a species and in combination with geologic history, they can improve our understanding of evolutionary processes (Lomolino et al. 2006; Riddle 1996). I then examine these concepts in more detail, first in the context of the diversity of fishes in the salmonid subfamily Coregoninae, then with a specific focus on least cisco (*Coregonus sardinella*), the focus of this study. Because the least cisco is broadly distributed across a large latitudinal gradient in Alaska, they are exposed to diverse environmental conditions. I describe the habitat characteristics by region and what is known about least cisco biology across a range of latitudes, with special emphasis on the higher latitude habitats where very little is known about least cisco populations. Because the least cisco has not been studied extensively, the information presented in this thesis provides a useful baseline of genetic relationships among populations across Alaska. As climate change alters Alaskan freshwater habitats and change connections and barriers to gene flow for freshwater fishes, future assessments of genetic relationships among populations can be compared to this baseline information, and this knowledge can guide future management decisions.

Genetics and geography

The objectives that motivate modern research in phylogeography and population genetics can be traced back to the earliest beginnings of biogeographic thought in the 19th century. Biogeography traces its roots to the studies of Alfred Russel Wallace and Charles Darwin and their theories of biological evolution. A concept underlying their theories is that global habitat variation shapes the distribution of species diversity along physical and ecological gradients (Darwin 1839; Wallace 1876). Since that time, patterns and mechanisms of species richness gradients have been eagerly investigated (Brown 1981; Hillebrand 2004; Hutchinson 1959; Levinton 1979; MacArthur et al. 1966;

MacArthur and Wilson 1963; Ricklefs 2004; Ricklefs 2006a; Ricklefs 2006b; Rohde 1992). Phylogeography was introduced by Avise in 1987 with the application of tools from molecular phylogenetics to biogeographic studies to uncover the temporal and spatial processes that shape the geographic distributions of genetic lineages of organisms (Avise 2000). In contrast to phylogeographic studies, the concepts and methods of population genetics research are often tailored for the study of genetic diversity at small geographic scales and evolutionary processes that affect populations over shorter time periods.

The study of genetic variation has helped ecologists address numerous questions about a species. For example, multi-locus allele frequency data can provide information about population of origin for individuals and how many distinct populations might be found in a region (Selkoe and Toonen 2006). Highly polymorphic sequence or microsatellite data can provide information about the recent evolutionary history of populations, such as recent changes in population size (Selkoe and Toonen 2006). Multi-locus genotype information can also provide information about genealogical relationships among individuals and whether individuals have moved (Selkoe and Toonen 2006). Other types of information ecologists may derive from the study of genetic variability include dispersal distance of offspring, source-sink relationships among populations, population structure and migration patterns in relation to the landscape, extinction/colonization dynamics, and connectivity among populations (Selkoe and Toonen 2006).

The influence of the glacial cycles and associated climatic fluctuations of the Pleistocene on the distribution of biological diversity in northwestern North American is the subject of a significant attention among phylogeographers (Shafer et al. 2010). Much of Canada was covered by ice-sheets during the Pleistocene. The largest recent ice-sheet coverage occurred approximately 20,000 years ago (Clark et al. 2009), and glacial advances and retreats, known as Croll-Milankovich cycles, occurred regularly (Imbrie 1985; Muller and MacDonald 1997). Populations of northwestern North American biota survived in glacial refugia and other suitable habitats (Holderegger and Thiel-Egenter

2009). The two major ice refugia were Beringia and the Pacific Northwest (Shafer et al. 2010). The Alexander Archipelago, Haida Gwaii, and the Canadian Arctic have been suggested as other important refugia (Macpherson 1965; Pielou 1991). Species also survived the ice ages by shifting their ranges across refugia. These groups were often isolated and genetically differentiated from conspecific populations over time, which shaped the surviving genetic diversity within species (Shafer et al. 2010). Today, species with broad geographic distributions are exposed to diverse environmental conditions across their range, each with a unique suite of environmental conditions to which organisms must adapt, leading to subpopulations that exhibit unique genetic characteristics (Witt et al. 2011).

Phylogeographic studies have demonstrated that Beringia was a refuge for numerous species, including: mountain avens, *Dryas integrifolia* (Tremblay and Schoen 1999), purple saxifrage, *Saxifraga oppositifolia* (Abbott and Comes 2004; Abbott et al. 2000), locoweeds, *Oxytropis* spp. (Jorgensen et al. 2003), white spruce (Anderson et al. 2006), Townsend's daisy, *Townsendia hookeri* (Thompson and Whitton 2006), lemmings, *Lemmus* (Fedorov et al. 2003), tundra voles, *Microtus oeconomus* (Brunhoff et al. 2003), thinhorn sheep, *Ovis dalli* (Loehr et al. 2006), collared pikas, *Ochotona collaris* (Hoberg et al. 2009), Alaska marmot, *Marmota broweri* (Steppan et al. 1999), Arctic shrew, *Sorex arcticus* (Fumagalli et al. 1999), and brown bears, *Ursus arctos* (Leonard et al. 2000). Phylogeographic studies of freshwater species have shown genetic substructuring within Beringia, supporting the refugia within refugia hypothesis. Breaks in ice-sheets and the Yukon River are the proposed barriers creating such phylogeographic breaks (Eddingsaas et al. 2004; Jorgenson et al. 2003). For example, a freshwater refugium of periglacial freshwater lakes likely existed for lake trout, *Salvelinus namaycush* (Wilson and Hebert 1998, and Arctic grayling, *Thymallus arcticus* (Stamford and Taylor 2004), in the southeastern portion of Beringia, at the juncture of the Cordilleran and Laurentide ice-sheets (Dyke and Prest 1987).

Freshwater fishes are model species in many studies linking patterns of genetic variation to geography because freshwater habitats offer limited dispersal opportunities to

the organisms that inhabit them (Lévêque et al. 2008). These constraints are partially responsible for creating the diversity in freshwater fishes seen around the world because populations may differentiate as a result of isolation or trough adaptation to local conditions (Lévêque et al. 2008). Phylogeographic studies offer the opportunity to better characterize the current distribution of biological diversity and to test hypothesis concerning the role of different historical and ecological factors in the evolution of that diversity.

Phylogeography

Phylogeography compares differences in genetic sequences to trace lines of descent back through time (Higgs and Attwood 2009). It explores the impacts of natural selection, population demography, and alternative historical scenarios on the spatial arrangement of genetic lineages (Avise 2000). It can also be used to identify historical hybridization events, hybrid zones, occurrences of introgression (Gonzalez-Rodriguez et al. 2004; Hewitt 2001; Swenson and Howard 2005), and geographic determinants of isolation. Phylogenies represent relationships among groups or taxa and are used in many branches of biology. They are models of inferred genealogical history and can describe: relationships among species on the tree of life; relationships between paralogues in gene families (Mäser et al. 2001); histories of populations (Edwards 2009); evolutionary and epidemiological dynamics of pathogens (Grenfell et al. 2004; Marra et al. 2003); genealogical relationships of somatic cells during differentiation and cancer development (Salipante and Horwitz 2006); evolution of language (Gray et al. 2009); and metagenomic sequences (Brady and Salzberg 2011). The integration of phylogeography with species range distribution models (i.e., ecological niche models) can help reveal how isolation, speciation, and selection are directly or indirectly linked to abiotic factors (Kozak et al. 2008).

Evolutionary forces such as recombination, mutation, and selection act differentially across genomes, according to the role of the gene (i.e. coding or noncoding) (Avise 2000; Fu and Li 1999). Evolutionary research relies on the concept that evolution imparts heterogeneity among the genomes of organisms within a particular taxon over

time (Sunnucks 2000). In animals, DNA is located in the nucleus and mitochondria (mtDNA). Phylogeographic studies can be designed to investigate sequences from either nuclear DNA or mtDNA, or both. Nuclear DNA is diploid and offspring inherit it from both parents. Some parts of the genome contain coding sequences, or genes that are transcribed to the amino acid sequences that make proteins, while other parts of the genome are not transcribed.

The mitochondrial genome is haploid and maternally inherited. It has a fast pace of sequence evolution because mutation rates are high and mtDNA is not constrained by histone proteins in its structure (Gillespie 1986; Li 1997; Nedbal and Flynn 1998; Richter 1992; Wilson et al. 1985). Most vertebrate mitochondrial genomes consist of 37 functionally distinct genes, which code for the control region (CR), 22 transfer RNAs, two ribosomal RNAs, and 13 messenger RNAs. The CR generally consists of three domains, a conserved central region flanked by two hypervariable domains on either end of the region that have high rates of nucleotide substitution and high levels of intraspecific polymorphisms (Lunt and Hyman 1997; McMillan and Palumbi 1997). The CR has been widely used in phylogeographic studies over microevolutionary scales of thousands to tens of thousands of years (Ward et al. 1993).

Phylogeographic surveys of mtDNA variation describe the current geographical structure of intraspecific matriarchal genealogies, and this information can yield evidence of past and present demographic factors that shape the geographic distribution of matrilines (Avise 2000). Variation in mtDNA is an informative measure for understanding population demography because: (i) females are generally philopatric to natal sites (Avise 2000); (ii) females and offspring generally remain in the same vicinity at the beginning of the offspring's life (Avise 2000); and (iii) strong matrilineal population structure implies demographic autonomy for local populations over shorter periods of time (Avise 2000).

Building a phylogeny from genetic information begins with aligning DNA sequences from currently living descendants. During alignment, bases are matched across sequences, and deletions and insertions are reflected by placing gaps in sequences. The

differences among the sequences in the alignment are compared and phylogenies are constructed to depict specific evolutionary relationships among taxa. The trees consist of nodes connected by branches, where branches depict the persistence of genetic lineages through time, and nodes denote the appearance of new lineages (Yang and Rannala 2012).

Phylogenies can be inferred by several methods that are either distance-based or character-based. For distance-based methods, the genetic distance (or degree of difference) is calculated between every pair of sequences in the data set. The resulting matrix of distances is then used for tree reconstruction. Character-based methods like maximum parsimony, maximum likelihood, and Bayesian analysis compare all sequences simultaneously to calculate a tree score. Theoretically, these methods can identify the tree with the best score by comparing all possible trees that can be created with the data (Yang and Rannala 2012).

Probabilistic models of DNA mutation

All methods except maximum parsimony use a probabilistic model of nucleotide substitution in their calculations (Yang and Rannala 2012). These models account for the knowledge that nucleotide sequences within the genome do not all change in the same way and at the same rate over evolutionary history. They also account for hidden substitutions even though they are not observed when sequences are compared. These models allow for a measure of total change, rather than just directly observed change, as a measure of dissimilarity. The foundation for all these models of mutation is the Markov model, in which the probabilities of state changes are independent of earlier history. Model variations include the Jukes-Cantor models (JC69 model) (Jukes and Cantor 1969), the Kimura models (K80 model) (Kimura 1980), HKY85 model (Hasegawa et al. 1985), and the General Time-Reversible model (GTR).

Maximum parsimony

When using maximum parsimony, the best tree to infer from the data is one requiring the fewest changes in states of characters. In other words, a ‘good’ tree describes evolutionary history with as little change as possible (Felsenstein 2004). Parsimony was originally used to analyze discrete morphological characters (Yang and

Rannala 2012) but was applied to molecular data during the late 1970s. Fitch (Fitch 1971) and Hartigan (Hartigan 1973) developed an algorithm for finding the minimum number of changes on a binary tree (and for reconstructing the ancestral states to achieve the minimum).

Distance

Distance-based methods provide a numerical measure of how similar two taxa are, with a distance of 0 indicating they are the same, and a larger number indicating some degree of difference. This is called the p-distance, or uncorrected distance, and several algorithms can calculate this value. A disadvantage of distance methods is that they do not use the full information in the data because sequences are transformed to a single measure. They are used primarily for quick initial explorations of data (Felsenstein 2004). The Unweighted Pair-Group Method with Arithmetic means (UPGMA) algorithm uses dissimilarity measures to guess plausible lengths of a metric tree and builds it. At each step, the two most similar taxa together are joined, placed equidistant from a common ancestor, and then collapsed into a single group. Each step reduces the size of the distance matrix by one group/taxon; eventually, all of the taxa are joined into a tree. UPGMA places all taxa at the same distance from the root. While this feature might be desirable if we believe a molecular clock underlies the data, in other situations it could be problematic (Felsenstein 2004). The Neighbor Joining algorithm (NJ) is the most commonly used distance method for tree reconstruction in phylogenetics (Saitou and Nei 1987). It uses a clustering algorithm, which assigns a set of individuals to groups (or clusters) so objects of the same cluster are more similar to each other than those from different clusters (Yang and Rannala 2012). A hierarchical cluster analysis can be agglomerative (starting with single elements and successively joining them into clusters) or divisive (starting with all objects and successively dividing them into partitions) (Yang and Rannala 2012). Various averaging processes performed at each step smooth out some errors in fit. The Minimum Evolution algorithm (ME) (Desper and Gascuel 2002) considers each possible topological tree in turn and uses least-squares minimization to

determine edge lengths that best fit dissimilarity data. It then chooses from metric trees the one with the minimum total of edge lengths (Desper and Gascuel 2002).

Maximum likelihood and Bayesian analysis

Maximum likelihood (ML) and Bayesian analysis (BA) are the two dominant statistical paradigms in common use for inference of phylogenetic trees. While profound philosophical differences distinguish these paradigms, both assume a probabilistic model of the evolution of sequences on trees and attempt to find tree(s) and model parameters most in accord with the data.

Maximum likelihood (ML) was developed by R. A. Fisher in the 1920s as a statistical methodology for estimating unknown parameters in a model, and for choosing parameters that make it most probable to produce the data collected (McAlpine et al. 1994). Maximum Likelihood tree estimation involves two optimization steps: (i) optimization of branch lengths to calculate a tree score for each candidate tree; and (ii) a search in tree space for the maximum likelihood tree (Yang and Rannala 2012). The tree (topology) is a model instead of a parameter, and branch lengths on a given tree and substitution parameters are parameters in the model (Yang and Rannala 2012). This process typically assumes a nucleotide substitution model. Using a common rate matrix and not needing to consider variable root locations reduces the number of parameters, making optimization more tractable (McAlpine et al. 1994). Maximum Likelihood has several important statistical properties. Estimators are statistically consistent; that is, as the number of trials increases to infinity, estimators converge to true parameters with minimal variance (McAlpine et al. 1994). A level of uncertainty exists at each branch within a phylogenetic tree. Statistical techniques such as bootstrapping, jackknifing, and randomization tests resample the information in a data set (in this case, genetic sequences) to infer the variability of parameters in the phylogenetic model (Felsenstein 2004). Bradley Efron introduced bootstrapping statistical methods in 1979, bootstrapping creates a data matrix by sampling whole characters from a set of *n* characters, with replacement, *n* times, resulting in a group of trees estimating the phylogeny for each bootstrap replicate (Felsenstein 2004). Branches in the trees are assigned a probability of

occurrence, accounting for the fact that some branches may have a length of zero or may not be present at all in some of the estimated trees (Felsenstein 2004). In the final consensus tree, we can be more confident that branches with high bootstrapping values more likely reflect true evolutionary relationships among taxa.

Bayesian analysis (BA) was introduced to molecular phylogenetics in the late 1990s (Li et al. 2000; Mau and Newton 1997; Rannala and Yang 1996; Yang and Rannala 1997). The goal of BA is not to infer a single best choice of parameters to fit data, but rather to associate to each possible choice of parameter values a probability of it being the one that produced the data (Yang and Rannala 2012). Analysis begins by specifying a prior distribution on parameters (Yang and Rannala 2012), and the probability measures associated with the prior distribution indicate support for the parameter values (Felsenstein 2004). Values near 1 indicate strong support, and those near 0 indicate essentially no support (Felsenstein 2004). The posterior distribution is the distribution of parameters (or models) that are conditional on the data, combining information in the prior distribution and in the data (likelihood) (Yang and Rannala 2012). Early BA methods assumed a molecular clock, but development of a more efficient Markov chain Monte Carlo (MCMC) method, which attempts to give a sample chosen from the distribution, eliminated the clock assumption allowing independent branch lengths on unrooted trees (Yang and Rannala 2012).

Population genetics

In contrast to phylogeography, the spatial scale and time frame of population genetics studies are much smaller because they aim to characterize more recent evolutionary processes. Levels of genetic differentiation calculated between populations over smaller geographic distances can identify isolation by distance patterns (O'Reilly et al. 2004; Ruzzante et al. 1999), clinal variation (Nielsen et al. 2004), fragmentation (Lemaire et al. 2005), hybridization (Gum et al. 2005), and cryptic speciation (Fillatre et al. 2003). Genetic techniques, when applied to fisheries management, have assisted managers to: (i) infer connectivity among stocks, dispersal capability of populations, and stock boundaries (Bohonak 1999; Kinlan and Gaines 2003; Watts et al. 2008; Watts and

Thorpe 2006); (ii) select brood stocks for aquaculture fisheries (Chistiakov et al. 2006); (iii) map economically important quantitative traits and identify genes responsible for these traits (Chistiakov et al. 2006); (iv) investigate patterns of gene flow across landscapes (Olsen et al. 2011); and (v) predict how habitat changes may affect genetic connectivity among populations (Olsen et al. 2011).

Because of the difference in spatial and temporal scales, population genetics studies investigate genetic loci that evolve more rapidly. Microsatellite loci are commonly used in population genetics studies. They are characterized by short combinations of simple nucleotide sequences that are repeated in tandem (Cruz et al. 2005), are abundant throughout the nuclear genome, are selectively neutral, and show high levels of allele polymorphism (Chistiakov et al. 2006). Microsatellites evolve in the genome: (i) as frameshift mutations through slipped-strand mispairing during DNA replication or repair (Kornberg et al. 1964); (ii) as interhelical junctions during chromosome alignment (Wilder and Hollocher 2001); (iii) as base substitutions (Wilder and Hollocher 2001); and (iv) as retrotransposition events (Wilder and Hollocher 2001). Their distribution across the genome is non-random (Katti et al. 2001) and their density varies across the genome (Lagercrantz et al. 1993). Additionally, microsatellite motifs, abundance, and mutation rates vary among species (Ross et al. 2003). Mutations of microsatellite markers may occur at a faster rate than in coding regions of the genome, estimated at 10^{-2}–10^{-6} per locus per generation (Ellegren 2000).

Microsatellite loci are amplified through targeted PCR reactions, and the sizes of the amplicons are measured to determine alleles. This process of determining alleles of a microsatellite is called scoring, and the combination of alleles from multiple loci is called a genotype. Scoring errors are a common concern when collecting microsatellite data, but no research conventions exist for reporting and mitigating errors (Bonin et al. 2004; Dakin and Avise 2004). The most common sources of error are stuttering, large-allele dropout, and null alleles. These potentially create consistent allelic and genotypic scoring bias by decreasing observed heterozygosity and skewing allele frequencies (DeWoody et al. 2006). Slipping by Taq polymerase during PCR can produce stutter bands around the

true allele (Jones and Avise 1997; Van Oosterhout et al. 2004). Interpreting stutter patterns can be particularly difficult for microsatellite loci with a two-nucleotide repeat motif, where true heterozygotes can be misinterpreted as homozygotes (at the larger allele). Large-allele dropout occurs during PCR, when the smaller allele of a true heterozygote is preferentially amplified, masking the larger allele (Björklund 2005). Finally, null alleles cause scoring errors because mutations at the priming site keep the allele from producing a visible amplification (Dakin and Avise 2004).

Quality control is an important factor in reporting accurate results for microsatellite data. Although a standard method for quality control has not been definitively established, some measures have been suggested for checking data quality. These include: (i) rescoring a subset of genotypes to calculate error rate and ensure accurate scoring (Selkoe and Toonen 2006); (ii) ensuring all alleles are amplified by testing for homozygote excess patterns consistent with null alleles (Selkoe and Toonen 2006); (iii) ensuring selective neutrality by testing for outlier loci (Selkoe and Toonen 2006); and (iv) performing defined crosses when possible to ensure Mendelian inheritance (Selkoe and Toonen 2006). Another means to improve data quality is to redesign primers for loci that commonly exhibit homozygote excess within populations.

When using microsatellites, statistical analyses must account for characteristics like mutation patterns. Because microsatellites are formed by repeating units of nucleotide base combinations, new mutant alleles involve a change in the number of repeat units. This means that the size of a new mutant allele depends on the size of its ancestor allele. The mutational process does not erase information about ancestral states (Slatkin 1995). Basic descriptive statistics of microsatellite variation are: number of alleles (N_A), expected heterozygosity (H_E), observed heterozygosity (H_O), allelic richness (A_R), which is included because the observed number of alleles per locus depends on sample size. F_{IS} is another descriptive statistic that can show the direction and magnitude of homozygote excess or deficit within a population.

Hardy-Weinberg equilibrium

The value for H_E is derived from Hardy-Weinberg equilibrium (HWE) equations using the allele frequencies given in the data. HWE equations are formulated to test for deviations from specific assumptions about the population being examined: random mating, normal Mendelian segregation of alleles, equal fertility of parent genotypes, equal fertilizing capacity of gametes, equal survival of all genotypes, a closed population (no migration), no mutation, a large population size (Frankham et al. 2002). Loci generally agree with these assumptions in large, naturally outbreeding populations, even if they are subject to evolutionary forces such as mutation, migrations, selection, and sampling effects (Frankham et al. 2002). H_E is then compared to H_O, which is the actual heterozygosity measured in a population, often with a $\boldsymbol{X}^2$ or pseudo exact test. Significant *p*-values indicate deviations from HWE assumptions.

***F*-statistics**

Sewall Wright developed a statistical method to measure the degree of differentiation among portions of population of organisms (Holsinger and Weir 2009). By assuming that differentiation among these portions, or sub-populations, was directly related to inbreeding within and among sub-populations, Wright used inbreeding coefficients (*F*) to describe the genetic diversity within and among sub-populations. In Wright's hierarchical model of *F*-statistics, he partitioned the inbreeding of individuals (*I*) in the total (*T*) population (F_{IT}), inbreeding of individuals relative to their sub-population (F_{IS}), and inbreeding of sub-populations relative to the total population (F_{ST}). F_{ST} is directly related to variance in allele frequency among populations and to the degree of resemblance among individuals within populations (Holsinger and Weir 2009). F_{ST} is low when gene flow rates are high among subpopulations or allele frequencies are similar among them, and F_{ST} is high when gene flow is limited and allele frequencies differ among subpopulations (Frankham et al. 2002; Holsinger and Weir 2009). More specifically, F_{ST} measures correlations due to common ancestry between random gametes from a subpopulation relative to random gametes from the total population (Wang 2012). The multiple interpretations of F_{ST} are further described in Table 1.1.

So many definitions of F_{ST} highlight the complexity of calculating and interpreting F_{ST} values that appropriately reflect levels of genetic differentiation among populations. Another factor to consider is that markers must be genetically neutral, meaning they are not found in the coding region of the genome or involve synonymous codons in a protein. All genetically neutral markers are expected to have the same F_{ST} value regardless of mutational patterns and rates, because mutations do not alter identity by descent (IBD) status (Wang 2012). Mutations do, however, obscure the genealogical history of genes because they change genes that had been identical in state (IIS) (Wang 2012). If certain pieces of information are available, such as pedigrees, that genealogical history can be recovered and provide more accurate estimates of F_{ST} when combined with genotype data (Wang 2012), but this is often not the case for natural populations.

Although F_{ST} should be an unbiased measure of genetic differentiation, factors such as the demographic history of a population (e.g. effective population size and migration rate), properties of the genetic markers (e.g. diversity and mutations), and sampling properties (e.g. sample size) can introduce bias (Wang 2012). The differences between genetic (or evolutionary) sampling and statistical sampling highlight how bias cannot be avoided. Genetic sampling results from genetic drift, a random process that happens in populations over time as a function of effective population size (N_e). Even if we went back in time and restarted the evolutionary process for a sampled population using the same known population sizes, mutation rates, migration rates, and selection coefficients for the current population, the genotype frequencies would usually differ because of the random nature of genetic drift (Rousset 2002). Statistical sampling is the variation associated with collecting genetic samples from a fixed set of population where genotype frequencies are unknown (Holsinger and Weir 2009). Increasing the size of within-population samples will reduce the amount of variation caused by statistical sampling. Additionally, *F*-statistics account for differences among sets of populations, but these sets are often only a subset of all the populations that exist or represent only one outcome of an underlying evolutionary process (Holsinger and Weir 2009).

Bias introduced by differences in sample sizes becomes apparent when PIBDs (probability of identity by descent), r_S and r_T, are considered. By definition (Wang 2012):

$$r_T = \sum_{i=1}^{S} \sum_{j=1}^{S} \sum_{k=1}^{N_i} \sum_{l=1}^{N_i} r_{ik,jl} / \sum_{i=1}^{S} \sum_{j=1}^{S} N_i N_j$$

$$r_S = \sum_{i=1}^{S} \sum_{j=1}^{N_i} \sum_{k=1}^{N_i} r_{ij,ik} / \sum_{i=1}^{S} N_i^2$$

The PIBD of the subpopulation (r_S) is a component of the PIBD of the total population (r_T), with the result that F_{ST} will be partially determined by the sample sizes of the subpopulations, with larger sample sizes having heavier weightings in the calculation of r_T (Wang 2012). Furthermore, r_T depends on *s*, or the number of subpopulations. The actual number of subpopulations in a natural population is often unknown. Because F_{ST} is estimated from a sample of subpopulations, it may not be truly reflective of the total population (Wang 2012).

F-statistics can also be limited for analyzing data from highly variable loci (Balloux et al. 2000; Balloux and Goudet 2002; Charlesworth 1998; Hedrick 1999; Jost 2008). Over time, other researchers have tried to address the limitations of F_{ST} by developing differentiation statistics that are conceptually similar to F_{ST}. These F_{ST} analogues include G_{ST} (Takahata and Nei 1984), θ (Weir and Cockerham 1984), R_{ST} (Slatkin 1995), Q_{ST} (Spitze 1993), Φ_{ST} (Excoffier et al. 1992), and *D* (Jost 2008). Each statistic employs specific modifications of the F_{ST} concept, although with their own limitations. For example, G_{ST} measures genetic differentiation due to all evolutionary forces, including genetic drift, migration, selection, and mutation (Wang 2012), and is appropriate only when the contribution of genetic drift to among population differences is not important (Holsinger and Weir 2009). G_{ST} is not a reliable measure of genetic differentiation when highly polymorphic markers are used (Wang 2012). R_{ST} is defined in terms of coalescent times, and gives unbiased estimates of differentiation due to demography (migration and population size). It is not affected by the amount of within-population variation, and was especially developed for markers with a stepwise mutation model, such as some microsatellites (Slatkin 1995). Mutation rates also do not affect R_{ST}, and it becomes more accurate with increasing mutation rates, because more mutations

allow a more accurate estimate of coalescent times (Slatkin 1995). However, R_{ST} is limited in that estimates are only satisfactory when mutations strictly follow the stepwise mutation model (Balloux et al. 2000). Φ_{ST} (for molecular sequence data) is useful when accounting for mutational 'distances' among alleles.

Debates about the advantages and disadvantages of using F_{ST} and its related analogue statistics are ongoing, with disagreements over the conditions in which each statistic is appropriate (e.g. genetic markers from different regions of the genome that may have varying mutational patterns and rates) (Holsinger and Weir 2009; Wang 2012). In a recent publication, Wang (2012) suggested that Wright's F_{ST} is a true measure of differentiation due to demographic factors only, whether it is defined in terms of inbreeding or coalescence time, and is independent of the type of genetic marker used.

The next step in population genetics analyses is to perform an AMOVA (Excoffier et al. 1992). F_{ST} values (or their analogs) are used in this method to analyze a hierarchical population structure, where individuals are grouped into populations and populations are grouped into higher-level clusters. The AMOVA then calculates within and between-groups sum of squares from a matrix of squared Euclidean distances between all pairs of individuals (Excoffier et al. 1992; Li 1976).

Beyond population genetics - metapopulation biology

Sometimes the *a priori* assumptions of what constitutes a population or subpopulation in a particular species may not be accurate, and the previously described statistical analyses may not be appropriate for describing genetic relationships among these groups. In these situations, it may be appropriate to view a group as a metapopulation instead of as a group of discrete populations. Metapopulation is a term used to describe any spatially structured population (Hanski 1998). The fundamental concepts of metapopulation biology were established by Levins (1969), in which he visualized a metapopulation as a 'population' consisting of a number of unstable local populations that inhabit discrete habitat patches, and a metapopulation persists as a balance between 'deaths' (local population extinctions) and 'births' (new populations established at unoccupied patches) (Levins 1969). Metapopulation biology examines the

dynamic consequences of migration and extinction/colonization processes among local populations of a species, with the landscape being a critical component affecting these processes (Hanski 1998). Metapopulation biology attempts to balance the theories of theoretical ecology, which concentrates on complex population dynamics and landscapes are considered homogeneous, and landscape ecology, emphasizing the structure of complex real landscapes with less attention paid to population dynamics (Hanski 1998). The compromise of metapopulation biology is that it views landscapes as networks of idealized habitat patches, and species occur as discrete local populations that are connected through migration (Hanski 1998).

Least cisco background information

Limited information exists on the biology of least cisco. Therefore, this section provides an overview of coregonin biology because much of it is expected to be applicable to our understanding of least cisco biology. This section also reviews what is known of coregonin evolutionary history, and the genetic data gathered in this project will help fill in the information gaps on least cisco. Additionally, this section highlights the importance of whitefish as a subsistence food for Alaskan Natives. Subsistence managers can make more informed management decisions as data from this and other projects become available.

Evolution of Salmonidae

Least cisco are in the Salmonidae family. Salmonids have been model species for systematic and phylogenetic research (Behnke 1972; Crespi and Fulton 2004; Norden 1961; Stearley and Smith 1993) because their life history characteristics offer an opportunity to explore evolutionary and ecological concepts such as mechanisms of speciation (Bernatchez 2004), the evolution of complex life histories (Crespi and Teo 2002), the role of hybridization in evolution (Taylor 2004), and patterns of chromosomal evolution (Phillips and Rab 2001). Salmonidae is characterized by a chromosome tetraploidization. In other words, the number of copies of some of their chromosomes doubled from two to four. Research suggests salmonids underwent a rapid radiation 25 to 100 million years ago following this tetraploidization event (Allendorf and Thorgaard

1984; Johnson et al. 1987). The family is broken down into three subfamilies: Coregoninae (ciscoes, whitefish, and inconnu), Thymallinae (grayling) and Salmoninae (huchen, lenok, trout, char, and salmon) (Kendall and Behnke 1984; Norden 1961; Stearley and Smith 1993). For decades, morphological and molecular data have suggested that Coregoninae is the sister group to the remainder of Salmonidae (Stearley and Smith 1993; Wilson and Li 1999). A recent phylogenetic study using a more comprehensive set of nuclear genes refuted this paradigm, suggesting Thymallinae is the sister group to the remainder of Salmonidae (Koop et al. 2008). Such competing lines of evidence highlight the fact that even with decades of research, questions about the evolutionary histories of salmonids remain (Crête-Lafrenière et al. 2012).

Evolutionary relationships among salmonids remain unresolved because of limitations from biological factors affecting salmonids, such as parallel and convergent evolution caused by species having similar ecological niches, rapid radiation, frequent hybridization, and local adaptation (Kinnison and Hendry 2004). Furthermore, evolutionary relationships remain unresolved because of a lack of more comprehensive genetic surveys (Bernatchez 2004; Koop et al. 2008). Although a recent comprehensive phylogenetic analysis of Salmonidae supports the monophyly of the Coregoninae subfamily (Crête-Lafrenière et al. 2012), lack of resolution is especially apparent within Coregoninae because of difficulty in species identification (Crête-Lafrenière et al. 2012), parallel evolution (Bernatchez 2004; Østbye et al. 2006), phenotypic plasticity (Lindsey 1981), recurrent trophic polymorphisms (Bernatchez et al. 1999; Douglas et al. 1999), contemporary hybridization (Taylor 2004), and historical introgression (Turgeon and Bernatchez 2003). The widely accepted morphological groupings of coregonids into whitefish and ciscoes may also not reflect evolutionary history (Bernatchez et al. 1991), in that recent analyses have shown that members of the sardine cisco clade (*C. sardinella, C. albula* and *C. peled*) are grouped with ‘true’ whitefish species (Crête-Lafrenière et al. 2012).

Evolutionary history and biology of *Coregonus*

Coregonins (whitefishes, ciscos and roundfish) have been model species from the onset of phylogeographic and population genetic studies, with such studies helping to advance and modify theories and methods in both (Bernatchez 2004; Bernatchez et al. 1999; Bernatchez et al. 2010). The northern aquatic ecosystems coregonins inhabit offer ideal conditions in which adaptive radiation can take place, which occurs when individuals of an ancestral lineage are exposed to differing environments with associated differences in selection pressures (Bernatchez 2004). Whitefishes are classified in the subfamily Coregoninae of the family Salmonidae. Coregoninae includes three genera: *Prosopium, Stenodus,* and *Coregonus*. *Prosopium* is characterized by a single nasal flap, parr mark retention, and the presence of a basibranchial plate, which is not present in *Coregonus* or *Stenodus* (Norden 1961).

The genus *Coregonus* is relatively diverse, with 28 species currently recognized, the highest number among salmonid genera (Reshetnikov 1988). However, studies of coregonin phylogenetic relationships indicate that greater than 28 species most likely exist (Bernatchez and Dodson 1994; Bodaly et al. 1991). Phenotypic diversity within and among species results in taxonomic confusion (Lindsey 1981). A survey of New and Old World coregonin fishes suggests that populations aggregate into six distinct genetic lineages: (i) Inconnu (genus *Stenodus*); (ii) an Arctic cisco group consisting of Arctic cisco (*C. autumnalis*) from the lower Mackenzie basin of Canada, widely disjunct Irish pollan (*C. pollan*), and lake cisco (*C. artedi*) populations sampled over a wide geographic range; (iii) a European (*C. lavaretus*) and lake whitefish (*C. clupeaformis*) group, including British schelley (*C. stigmaticus*) and powan (*C. clupeoides*), and lake whitefish from North America that clustered in two distinct groups; (iv) broad whitefish (*C. nasus*) that constituted a relatively homogeneous group, but were more closely related to European and lake whitefish than other coregonin groups; (v) a least cisco (*C. sardinella*) and vendace (*C. albula*) group; and (vi) a peled (*C. peled*) group that formed a separate genetic grouping but were more closely related to vendace and least cisco than other coregonins (Bodaly et al. 1991).

Generally, whitefishes require cold, well-oxygenated aquatic habitats (Scott and Crossman 1973). Ciscoes often inhabit pelagic lacustrine zones and feed on zooplankton (Scott and Crossman 1973), while true whitefish generally inhabit the epibenthic zone (Lindsey 1981) and vary in movement patterns. For example, Arctic (*C. autumnalis*) and Bering cisco (*C. laurettae*) are obligate anadromous species, and migrate hundreds of kilometers upstream to spawn (McPhail and Lindsey 1970). Other species, including those from the *C. clupeaformis/lavaretus* species complex, live out their life cycles in river and lake habitats, but also can have populations with anadromous life histories.

A life history characteristic generally applicable to riverine whitefish populations is that they are composed of three main groups: immature fish located downstream of spawning areas; mature non-spawning fish located downstream of spawning areas but not necessarily in the same place as immature fish; and mature spawners located at or near upstream spawning areas (Reist and Bond 1988). Whitefish may not feed during their upstream migration to spawning sites (Alt 1969), although feeding behavior data from the Selawik River drainage suggest this may vary by species (Brown 2004). Whitefish exhibit high fidelity to natal spawning sites (Hallberg 1989). It can take between four and eight years for whitefish to mature and be ready for spawning migration (Morrow 1980).

Least cisco

Little is known of the genetic relationships among populations of least cisco (*C. sardinella*) throughout their range, and almost no research has been done on Alaskan populations of least cisco (Dillinger 1989). Least cisco likely survived through quaternary glaciation events in two separate refugia, and the descendants of these two groups have since become sympatric (Dillinger 1989). Many studies have focused on other whitefish species in Alaska because they are important targets of commercial and subsistence fisheries. Yet, as concerns grow over the health and size of all whitefish populations, more attention may turn to less frequently targeted species like least cisco to increase harvest sizes and provide year-to-year stability to subsistence catches.

Least cisco range across Arctic Canada, Alaska, and Siberia, encountering highly variable environmental conditions across a broad range of spatial and temporal scales. In

Canada, they are distributed in the southern portion of the Mackenzie River to Fort Simpson, and extend east to Bathurst Inlet and Cambridge Bay (Mecklenburg et al. 2002). Further west in Alaska, they are distributed from the Arctic Coast to Bristol Bay, inhabiting streams and lakes north of the Alaska Range and the drainages of the Yukon and Kuskokwim Rivers. The most western portion of their range extends into Siberia and the White Sea (Mecklenburg et al. 2002). Least cisco are broadly distributed across freshwater drainages in Alaska, and often share habitat with other whitefish species. Because least cisco are so broadly distributed across Alaska, populations in different geographic locations will experience unique suites of environmental conditions. Consequently, life history strategies and phenotypic adaptations may not be uniform across the species' range. Therefore, habitat variation is an important component that shapes life history and phenotypic variations within the species.

The life history variation and present-day distribution of least cisco suggest that a combination of phylogeny and physiological responses to the environment has equipped this species with the ability to adapt to extreme and variable environmental conditions in comparison to other whitefish species (Dillinger 1989). Least cisco life history strategies vary by latitude (Dillinger 1989). Those in more northern populations in the Mackenzie River reach maturity at between six and eight years (Mann 1974), and spawning may occur in alternate years or less frequently (Dillinger 1989). Conversely, least cisco populations further south in the Yukon River region mature at approximately four years, and most appear to spawn annually (Dillinger 1989). These differences may result from differences in access to resources. Energy resources may become scarcer with increasing latitude, resulting in slower growth and reduced spawning frequency (Dillinger 1989). Additionally, life history variation may also result from differences in behavioral and phenotypic traits. For example, six least cisco populations along the Beaufort Sea coast exhibit phenotypic variation between individuals co-occurring in a lake (Dillinger 1989; Mann 1974). Additionally, populations of both normal and dwarf size least cisco inhabit Trout Lake in the Yukon Territory of Canada and Peters Lake, and all were freshwater resident. Conversely, the anadromous riverine least cisco from the Mackenzie River are

larger, heavier, longer-lived, and more fecund, on average, than the least cisco from the lakes (Dillinger 1989; Mann 1974; Mann and McCart 1981).

Further examples of least cisco life history variation are found across Alaska. In the Kuskokwim River drainage, the second largest drainage in Alaska, least cisco enter freshwater tundra ponds and lakes in the early spring because the amount of oxygen has risen to a tolerable level, and remain in these ponds and lakes to feed throughout the spring and summer (Alt 1979). This movement into the shallow tundra ponds and lakes in the early spring appears to be a critical component in their life history. They likely do not use these lakes in the winter because dissolved oxygen levels are too low (Harper et al. 2007). Movements in Whitefish Lake, which drains into the upper end of the Lower Kuskokwim drainage, occur very early in the spring. The highest numbers of least cisco emigrate from Whitefish Lake at the end of July, but other emigration pulses also occur toward the end of August, the middle of September, and the beginning of October. Variation in emigration timing between years may be the consequence of needing additional forage time to replenish fat reserves (Harper et al. 2007). Some fish remain in the lake for two months while others remain for six. Some fish from Whitefish Lake move to the upstream region of the Kuskokwim River, while others move downstream as far as Tuluksak (Harper et al. 2007).

Further north, least cisco inhabit the upper Koyukuk River basin with inconnu (*S. leucichthys*), broad whitefish (*C. nasus*), humpback whitefish (*C. clupeaformis*), and round whitefish (*P. cylindraceum*). Least cisco from the upper Koyukuk River move widely throughout the drainage, and have been tracked in the Koyukuk, Alatna, Kanuti, and South Fork Rivers (Brown 2009), with spawning habitat in the Alatna and Kanuti Rivers. Least cisco also co-inhabit the Selawik River with the same suite of whitefishes as in the Koyukuk. Many least cisco here appear to remain in freshwater throughout their lives. They mature by age five, but the Selawik River drainage does not include habitats with flowing water and gravel substrate, which is the preferred spawning habitat for whitefish. Therefore, least cisco in the Selawik River drainage likely migrate long distances to find suitable spawning habitat (Brown 2004).

Finally, least cisco populations have been documented in many locations along the northern coast of Alaska, including the Colville River, Camden Bay, Beaufort Lagoon, and Barter Island (Brown 2008). Little is known about the least cisco populations inhabiting the inland freshwater lakes of the Arctic Coastal Plain (Brown 2004; Brown 2008; Brown 2009; Brown et al. 2007; Dillinger 1989; Harper and Harris 2006; Harper et al. 2007; Harper et al. 2009; Sutton and Edenfield 2012), but the few studies available do reveal that least cisco populations exhibit high levels of phenotypic and life history plasticity. For example, they can be amphidromous, lacustrine, or riverine (Harper et al. 2007). These populations have likely been overlooked because they are not subject to intense subsistence harvests since the lakes are very remote and difficult to access.

Whitefishes and Native Alaskan subsistence

In Alaska, whitefishes are the most abundant group of fish north of the Alaska Range, and are located in all freshwater habitat types (Alt 1994). They are an important food source to predatory fish and piscivorous birds on the North Slope. Additionally, whitefishes are a valued subsistence food in rural Alaskan communities, and are harvested at a few small commercial fishing operations (Alt 1994; Carter 2010). The Arctic cisco (*Coregonus autumnalis*) is a highly valued food resource (Moulton et al. 2010) in the Colville River drainage area, and anadromous least cisco are the primary by-catch at the fall fisheries. Least cisco are of special importance in years when Arctic cisco harvests are low. They are also shared with other communities or used as dog food (Moulton et al. 2010). The season for catching fish occurs in the late summer near Kaktovik and after the fall ice formation in the Colville River Delta near Nuiqsut (Moulton et al. 2010). After ice forms on the river delta, fishermen string gill nets under the ice to catch cisco. Least cisco have comprised as little as 3.8% of the total catch in 1986, to as much as 50.8% of the total catch in 1998 (Seigle and Parrett 2008).

The Selawik and Kuskokwim River drainages are also sites for whitefish subsistence fishing (Brown 2004), and overexploitation or loss of the fishery could dramatically alter Alaskan Natives' local economies (Harper et al. 2009). Methods of harvest in Whitefish Lake drainage include gillnets under the ice in winter, gillnets in the

spring, summer, and fall, spear fishing in fall and winter, and hook and line (Simon et al. 2007). Intensive whitefish subsistence harvests occur in small areas over many hundreds of kilometers of the Kuskokwim River, and catches from these harvests comprise mixtures of locally available whitefish species. Consequently, this complicated harvest system may put weaker fish stocks at risk of overexploitation.

Recently, the Iñupiat of Nuiqsut have expressed concerns over reduced availability of fish, reduced fish size, weight and quality, changes in distribution, deformities, changes in food web interactions, and stress to the fish caused by development obstructions from the oil industry (Murphy et al. 2007). More biological and ecological information on the whitefish populations across Alaska is required to improve management plans. Managers have been working in the Kuskokwim River drainage to better understand the migratory behavior, growth, population size, survival rates, fine-scale movements, and fidelity to spawning and feeding grounds of whitefish species (Harper and Harris 2006). However, whitefish fisheries extend beyond the Kuskokwim, so research must go beyond this region as well to gather more information about Alaska's whitefishes.

Climate change and least cisco in the Arctic

The central concept of both phylogeographic and population genetic studies is that temporal, spatial, and ecological processes shape the geographic distributions of genetic lineages of organisms. Arctic aquatic systems are expected to act as key indicators of the timing, rate, intensity, and effects of the change. The Arctic may be particularly susceptible to climate change for a number of reasons: (i) positive feedback effects of greater heat absorption associated with lower snow and ice cover on land and sea, (ii) larger fraction of energy going to warming rather than evaporation relative to the tropics, (iii) shallower troposphere (lower atmosphere) and frequent surface-based temperature inversions, and (iv) atmospheric and oceanic circulation (ACIA 2004). This region is warming at a rate almost twice the global average, with the most pronounced warming occurring during winter and spring (ACIA 2004), resulting in changes such as rising ocean temperatures, sea level rise, rapidly eroding shorelines, increased storm

surges, melting ground ice, permafrost degradation, changes to river discharge and sediment transport, and increased shrub growth at high latitudes.

The ultimate ecological effects of climate change are not fully understood, but projections predict that water temperatures will increase beyond what is optimum for fish, resulting in decreased biomass and yield, altered rates, and locations of colonization, extinction, competition, productivity (Prowse et al. 2006a; Prowse et al. 2006b; Prowse et al. 2006c; Tonn 1990), and available habitat, because warmer waters will alter the thermocline in lakes, reducing available habitat for fish adapted to colder water below these thermoclines (Reist et al. 2006a). Additionally, higher air temperatures and longer ice-free seasons may cause thermal stratification in lakes that currently remain unstratified due to wind-mixing, short ice-free seasons, and low air temperatures (Martin et al. 2009). Habitat alterations will have large impacts on Arctic wildlife and necessitate management plans that take these changes into account (Martin et al. 2009). Although the responses of Arctic wildlife to climate change are poorly understood (ACIA 2004), it is widely accepted that climate change will have widespread consequences on the physical, chemical, and biological processes that shape fish habitat in the Arctic (Reist et al. 2006a; Reist et al. 2006b).

Phylogeographic and population genetic investigations will provide insight into how highly mobile freshwater fishes like least cisco use connections within hydrographic networks, or are blocked by barriers in the landscape. The population genetics portion of this study focuses on least cisco collected on the Arctic Coastal Plain. We not only lack information for least cisco from these populations overall, but building a better understanding of how freshwater fish populations are genetically structured across the Alaskan landscape is particularly critical for understanding the implications of dynamic present-day conditions. Water connectivity and the barriers created by landscape features like mountain ranges play an important role in gene flow among populations. As climate change affects the hydrographic networks across Alaska, particularly the lake systems on the North Slope, the connections and barriers that shape gene flow may change. This study can provide baseline genetic information that can serve as a means of comparison

in the future, when changes in genetic relationships may result from landscape alterations caused by climate change.

Project overview

The questions guiding my thesis concern the genetic relationships among least cisco on different landscape scales in Alaska. First, on the geographic scale of the whole state, how are northern least cisco populations genetically related to southern least cisco populations? Second, how are least cisco genetically related to each other on a finer landscape scale, such as among a group of closely located freshwater lakes on the Arctic Coastal Plain? Therefore, my objectives are: (i) to assess levels of genetic variability in and patterns of genetic connectivity among Alaskan populations of least cisco; and (ii) to characterize fine-scale population genetic structure of least cisco from the Arctic Coastal Plain. To accomplish the first objective, I compared DNA sequences from the control region (CR) of the mitochondrial genome (mtDNA) of least cisco from the Arctic Coastal Plain (ACP), Barrow (BAR), Point Lay (LAY), Selawik River (SEL), Koyukuk River (KOY), Rampart Rapids (RAP), Whitefish Lake (WHI), Lake Clark (CLA), and Toksook River (TOK). I examined the diversity and geographic distribution of sampled CR haplotypes to identify phylogeographic breaks and estimate the degree of connectivity among these putative populations. To accomplish the second objective, I genotyped nuclear genome microsatellite loci in least cisco collected from eight freshwater lakes on the Arctic Coastal Plain. I examined the diversity of microsatellite loci and compared levels of diversity to estimate the degree of genetic differentiation among lakes.

References

Abbott, R.J., and Comes, H.P. 2004. Evolution in the Arctic: a phylogeographic analysis of the circumarctic plant, *Saxifraga oppositifolia* (Purple saxifrage). New. Phytol. **161**(1): 211-224.

Abbott, R.J., Smith, L.C., Milne, R.I., Crawford, R.M.M., Wolff, K., and Balfour, J. 2000. Molecular analysis of plant migration and refugia in the Arctic. Science **289**(5483): 1343-1346.

ACIA. 2004. Impacts of a Warming Arctic: Arctic Climate Impact Assessment. Cambridge University Press, New York. 144pp.

Allendorf, F.W., and Thorgaard, G.H. 1984. Tetraploidy and the evolution of salmonid fishes. Evol. Genet. Fish. Monogr. Evol. Biol. pp. 1-53.

Alt, K. 1994. Whitefish Species: Wildlife Notebook Series. Alaska Department of Fish and Game. 1p.

Alt, K.T. 1969. Taxonomy and ecology of the inconnu, (*Stenodus leucichthys* nelma), in Alaska. Biological Papers of the University of Alaska **12**: 61pp.

Alt, K.T. 1979. A life history and study of sheefish and whitefish in Alaska. Annual Performance Report, 1980-1981, Federal Aid in Fish Restoration, Project F-9-11, Vol. 20, Juneau, AK. 19pp.

Anderson, L.L., Hu, F.S., Nelson, D.M., Petit, R.J., and Paige, K.N. 2006. Ice-age endurance: DNA evidence of a white spruce refugium in Alaska. Proc. Natl. Acad. Sci. U.S.A. **103**(33): 12447-12450.

Avise, J.C. 2000. Phylogeography: The History and Formation of Species. Harvard University Press, Cambridge, MA. 447pp.

Balloux, F., Brunner, H., Lugon-Moulin, N., Hausser, J., and Goudet, J. 2000. Microsatellites can be misleading: an empirical and simulation study. Evolution **54**: 1414-1422.

Balloux, F., and Goudet, J. 2002. Statistical properties of population differentiation estimators under stepwise mutation in a finite island model. Mol. Ecol. **11**(4): 771-783.

Behnke, R.J. 1972. The systematics of salmonid fishes of recently glaciated lakes. J. Fish. Res. Board Can. **29**(6): 639-671.

Bernatchez, L. 2004. Ecological theory of adaptive radiation: an empirical assessment from coregonine fishes (Salmoniformes). *In* Evolution illuminated: salmon and their relatives. *Edited by* A.P. Hendry and S.C. Stearns. Oxford University Press, Oxford, UK. pp. 175-207.

Bernatchez, L., Colombani, F., and Dodson, J.J. 1991. Phylogenetic relationships among the subfamily Coregoninae as revealed by mitochondrial DNA restriction analysis. J. Fish Biol. **39**(A): 283-290.

Bernatchez, L., Chouinard, A., and Lu, G.Q. 1999. Integrating molecular genetics and ecology in studies of adaptive radiation: whitefish, *Coregonus* sp., as a case study. Biol. J. Linn. Soc. **68**(1-2): 173-194.

Bernatchez, L., and Dodson, J.J. 1994. Phylogenetic relationships among palearctic and nearctic whitefish (*Coregonus* sp) populations as revealed by mitochondrial DNA variation. Can. J. Fish. Aquat. Sci. **51**(1): 240-251.

Bernatchez, L., Renaut, S., Whiteley, A.R., Derome, N., Jeukens, J., Landry, L., Lu, G., Nolte, A.W., Ostbye, K., Rogers, S.M., and St-Cyr, J. 2010. On the origin of species: insights from the ecological genomics of lake whitefish. Philos. T. R. Soc. B **365**(1547): 1783-1800.

Björklund, M. 2005. A method for adjusting allele frequencies in the case of microsatellite allele drop-out. Mol. Ecol. Notes **5**(3): 676-679.

Bodaly, R.A., Vuorinen, J., Wards, R.D., Luczynski, M., and Reist, J.D. 1991. Genetic comparisons of New and Old-World coregonid fishes. J. Fish Biol. **38**(1): 37-51.

Bohonak, A.J. 1999. Dispersal, gene flow, and population structure. Q. Rev. Biol. **74**(1): 21-45.

Bonin, A., Bellemain, E., Eidesen, P.B., Pompanon, F., Brochmann, C., and Taberlet, P. 2004. How to track and assess genotyping errors in population genetics studies. Mol. Ecol. **13**: 3261-3273.

Brady, A., and Salzberg, S. 2011. PhymmBL expanded: confidence scores, custom databases, parallelization and more. Nat. Methods **8**: 367.

Brown, J.H. 1981. Two decades of homage to Santa Rosalia: Toward a general theory of diversity. Am. Zool. **21**(4): 877-888.

Brown, R.J. 2004. A biological assessment of whitefish species harvested during the spring and fall in the Selawik River Delta, Selawik National Wildlife Refuge, Alaska: Alaska Fisheries Technical Report Number 77. U.S. Fish and Wildlife Service, Fairbanks Fish and Wildlife Field Office. 50pp.

Brown, R.J. 2008. Life History and Demographic Characteristics of Arctic Cisco, Dolly Varden, and Other Fish Species in the Barter Island Region of Northern Alaska. Alaska Fisheries Technical Report Number 101, November 2008, Fairbanks, AK. 52pp.

Brown, R.J. 2009. Distributions and demographics of whitefish species in the Upper Koyukuk River Drainage, Alaska, with emphasis on seasonal migrations and important habitats of broad whitefish and humpback whitefish: Alaska Fisheries Technical Report Number 104. U.S. Fish and Wildlife Service, Fairbanks Fish and Wildlife Field Office. 24pp.

Brown, R.J., Bickford, N., and Severin, K. 2007. Otolith trace element chemistry as an indicator of anadromy in Yukon River drainage coregonine fishes. Trans. Am. Fish. Soc. **136**(3): 678-690.

Brunhoff, C., Galbreath, K.E., Fedorov, V.B., Cook, J.A., and Jaarola, M. 2003. Holarctic phylogeography of the root vole (*Microtus oeconomus*): implications for late Quaternary biogeography of high latitudes. Mol. Ecol. **12**(4): 957-968.

Carter, W.K. 2010. Life history and spawning movements of broad whitefish in the Middle Yukon River, School of Fisheries and Ocean Sciences, University of Alaska Fairbanks, Fairbanks, AK. 55pp.

Charlesworth, B. 1998. Measures of divergence between populations and the effect of forces that reduce variability. Mol. Biol. Evol. **15**: 538-543.

Chistiakov, D.A., Hellemans, B., and Volckaert, F.A.M. 2006. Microsatellites and their genomic distribution, evolution, function and applications: A review with special reference to fish genetics. Aquaculture **255**(14): 1-29.

Clark, P.U., Dyke, A.S., Shakun, J.D., Carlson, A.E., Clark, J., Wohlfarth, B., Mitrovica, J.X., Hostetler, S.W., and McCabe, A.M. 2009. The last glacial maximum. Science **325**(5941): 710-714.

Crespi, B.J., and Fulton, M.J. 2004. Molecular systematics of Salmonidae: combined nuclear data yields a robust phylogeny. Mol. Phylogenet. Evol. **31**(2): 658-679.

Crespi, B.J., and Teo, R. 2002. Comparative phylogenetic analysis of the evolution of semelparity and life history in salmonid fishes. Evolution **56**(5): 1008-1020.

Crête-Lafrenière, A., Weir, L.K., and Bernatchez, L. 2012. Framing the salmonidae family phylogenetic portrait: A more complete picture from increased taxon sampling. Plos One **7**(10): e46662.

Crow J.F., and Kimura, M. 1970. An Introduction to Population Genetics Theory. Harper and Row, New York.

Cruz, F., Perez, M., and Presa, P. 2005. Distribution and abundance of microsatellites in the genome of bivalves. Gene **346**: 241-247.

Dakin, E.E., and Avise, J.C. 2004. Microsatellite null alleles in parentage analysis. Heredity **93**: 504-509.

Darwin, C. 1839. Narrative of the surveying voyages of His Majesty's Ships Adventure and Beagle between the years 1826 and 1836, describing their examination of the southern shores of South America, and the Beagle's circumnavigation of the globe. Journal and remarks. 1832-1836 III. Henry Colburn, London.

Desper, R., and Gascuel, O. 2002. Fast and accurate phylogeny reconstruction algorithms based on the minimum-evolution principle. J. Comput. Biol. **9**: 687-705.

DeWoody, J.A., Nason, J.D., and Hipkins, V.D. 2006. Mitigating scoring errors in microsatellite data from wild populations. Mol. Ecol. Notes **6**(4): 951-957.

Dillinger, R.E. 1989. An Analysis of Taxonomic Status of the *Coregonus autumnalis* Species Complex in North America, and an Investigation of the Life History of Whitefishes and Ciscoes (Pisces; Coregoninae) in North America and Eurasia, Memorial University of Newfoundland, St. John's, Newfoundland, Canada. 163pp.

Douglas, M., Brunner, P., and Bernatchez, L. 1999. Do assemblages of *Coregonus* (Teleostei : Salmoniformes) in the Central Alpine region of Europe represent species flocks? Mol. Ecol. **8**(4): 589-603.

Dyke, A.S., and Prest, V.K. 1987. The Late Wisconsinan and Holocene history of the Laurentide Ice Sheet. Geogr. Phys. Quatern. **41**: 237-263.

Eddingsaas, A.A., Jacobsen, B.K., Lessa, E.P., and Cook, J.A. 2004. Evolutionary history of the Arctic ground squirrel (*Spermophilus parryii*) in Nearctic Beringia. J. Mammal. **85**(4): 601-610.

Edwards, S.V. 2009. Is a new and general theory of molecular systematics emerging? Evolution **63**(1): 1-19.

Ellegren, H. 2000. Microsatellite mutations in the germline: implications for evolutionary inference. Trends Genet. **16**(12): 551-558.

Excoffier, L., Smouse, P.E., and Quattro, J.M. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes - Application to human mitochondrial-DNA restriction data. Genetics **131**(2): 479-491.

Falconer, D.S., and Mackay, T.F.C. 1996. Introduction to Quantitative Genetics. Prentice Hall, New York. 480pp.

Fedorov, V.B., Goropashnaya, A.V., Jaarola, M., and Cook, J.A. 2003. Phylogeography of lemmings (*Lemmus*): no evidence for postglacial colonization of Arctic from the Beringian refugium. Mol. Ecol. **12**(3): 725-731.

Felsenstein, J. 2004. Inferring Phylogenies. Sinauer Associates, Inc., Sunderland, MA. 664pp.

Fillatre, E.K., Etherton, P., and Heath, D.D. 2003. Bimodal run distribution in a northern population of sockeye salmon (*Oncorhynchus nerka*): life history and genetic analysis on a temporal scale. Mol. Ecol. **12**(7): 1793-1805.

Fitch, W.M. 1971. Toward defining the course of evolution: minimum change for a specific tree topology. Syst. Zool. **20**: 406-416.

Frankham, R., Ballou, J.D., and Briscoe, D.A. 2002. Introduction to conservation genetics. Cambridge University Press, Cambridge. 617pp.

Fu, Y.X., and Li, W.H. 1999. Coalescing into the 21st century: An overview and prospects of coalescent theory. Theor. Popul. Biol. **56**(1): 1-10.

Fumagalli, L., Taberlet, P., Stewart, D.T., Gielly, L., Hausser, J., and Vogel, P. 1999. Molecular phylogeny and evolution of Sorex shrews (Soricidae : Insectivora) inferred from mitochondrial DNA sequence data. Mol. Phylogenet. Evol. **11**(2): 222-235.

Gillespie, J.H. 1986. Variability of evolutionary rates of DNA. Genetics **113**(4): 1077-1091.

Gonzalez-Rodriguez, A., Bain, J.F., Golden, J.L., and Oyama, K. 2004. Chloroplast DNA variation in the *Quercus affinis-Q. laurina* complex in Mexico: geographical structure and associations with nuclear and morphological variation. Mol. Ecol. **13**(11): 3467-3476.

Gray, R.D., Drummond, A.J., and Greenhill, S.J. 2009. Language phylogenies reveal expansion pulses and pauses in pacific settlement. Science **323**: 479-483.

Grenfell, B.T., Pybus, O.G., Gog, J.R., Wood, J.L.N., Daly, J.M., Mumford, J.A., and Holmes, E.C. 2004. Unifying the epidemiological and evolutionary dynamics of pathogens. Science **303**: 327-332.

Gum, B., Gross, R., and Kuehn, R. 2005. Mitochondrial and nuclear DNA phylogeography of European grayling (*Thymallus thymallus*): evidence for secondary contact zones in central Europe. Mol. Ecol. **14**(6): 1707-1725.

Hallberg, J.E. 1989. Abundance and size composition of Chatanika River least cisco and humpback whitefish with estimates of exploitation by recreational fishermen: Fishery Data Series Report Number 108. Alaska Department of Fish and Game. 22pp.

Hanski, I. 1998. Metapopulation dynamics. Nature **396**(6706): 41-49.

Harper, K.C., and Harris, F. 2006. Pilot study, using PIT tags to monitor movements of Kuskokwim River whitefish. Alaska Fisheries Data Series Number 2006-14. U.S. Fish and Wildlife Service, Fairbanks Fish and Wildlife Field Office. 16pp.

Harper, K.C., Harris, F., Brown, R.J., Wyatt, T., and Cannon, D. 2007. Stock assessment of broad whitefish, humpback whitefish and least cisco in Whitefish Lake, Yukon Delta National Wildlife Refuge, Alaska, 2001-2003. Alaska Fisheries Technical Report Number 88. U.S. Fish and Wildlife Service, Fairbanks Fish and Wildlife Field Office. 55pp.

Harper, K.C., Harris, F., Miller, S.J., and Orabutt, D. 2009. Migration, timing and seasonal distribution of broad whitefish, humpback whitefish, and least cisco from Whitefish Lake and the Kuskokwim River, Alaska, 2004 and 2005. Alaska Fisheries Technical Report Number 105. U.S. Fish and Wildlife Service, Kenai, Alaska. 34pp.

Hartigan, J.A. 1973. Minimum evolution fits to a given tree. Biometrics **29**: 53-65.

Hasegawa, M., Kishino, H., and Yano, T. 1985. Dating the human-ape splitting by a molecular clock of mitochondrial DNA. J. Mol. Evol. **22**: 160-174.

Hedrick, P.W. 1999. Perspective: Highly variable loci and their interpretation in evolution and conservation. Evolution **53**(2): 313-318.

Hewitt, G.M. 2001. Speciation, hybrid zones and phylogeography - or seeing genes in space and time. Mol. Ecol. **10**(3): 537-549.

Higgs, P.G., and Attwood, T. 2009. Bioinformatics and Molecular Evolution. Wiley-Blackwell, Hoboken, NJ, USA. 384pp.

Hillebrand, H. 2004. On the generality of the latitudinal diversity gradient. Am. Nat. **163**(2): 192-211.

Hoberg, E.P., Pilitt, P.A., and Galbreath, K.E. 2009. Why Museums Matter: A Tale of Pinworms (Oxyuroidea: Heteroxynematidae) Among Pikas (*Ochotona princeps* and *O. collaris*) in the American West. J. Parasitol. **95**(2): 490-501.

Holderegger, R., and Thiel-Egenter, C. 2009. A discussion of different types of glacial refugia used in mountain biogeography and phylogeography. J. Biogeogr. **36**(3): 476-480.

Holsinger, K.E., and Weir, B.S. 2009. Fundamental concepts in genetics, genetics in geographically structured populations: defining, estimating and interpreting F_{ST}. Nat. Rev. Genet. **10**(9): 639-650.

Hutchinson, G.E. 1959. Homage to Santa Rosalia or Why are there so many kinds of animals? Am. Nat. **93**: 145-159.

Imbrie, J. 1985. A theoretical framework for the Pleistocene ice ages. J. Geol. Soc. **142**: 417-432.

Johnson, K.R., Wright, J.E., and May, B. 1987. Linkage relationships reflecting Ancestral tetraploidy in salmonid fish. Genetics **116**(4): 579-591.

Jones, A.G., and Avise, J.C. 1997. Microsatellite analysis of maternity and the mating system in the Gulf pipefish *Syngnathus scovelli*, a species with male pregnancy and sex-role reversal. Mol. Ecol. **6**(3): 203-213.

Jorgensen, J.L., Stehlik, I., Brochmann, C., and Conti, E. 2003. Implications of ITS sequences and RAPD markers for the taxonomy and biogeography of the *Oxytropis campestris* and *O. arctica* (Fabaceae) complexes in Alaska. Am. J. Bot. **90**(10): 1470-1480.

Jorgenson, M.T., Roth, J.E., Schlentner, S.F., and Cater, T.C. 2003. Ecological land evaluation for the Yukon Training Area on Fort Wainwright: permafrost and disturbance. Available from ABR, Fairbanks, Alaska, for U.S. Army, Anchorage, Alaska. 88pp.

Jost, L. 2008. G_{ST} and its relatives do not measure differentiation. Mol. Ecol. **17**: 4015-4026.

Jukes, T.H., and Cantor, C.R. 1969. Evolution of Protein Molecules. New York: Academic Press. pp. 21-132.

Katti, M.V., Ranjekar, P.K., and Gupta, V.S. 2001. Differential distribution of simple sequence repeats in eukaryotic genome sequences. Mol. Biol. Evol. **18**(7): 1161-1167.

Kendall, A.W., and Behnke, R.J. 1984. Salmonidae: Development and relationships. *In* Ontogeny and systematics of fishes. *Edited by* W.J.R.H.G. Moser, D.M. Cohen and M.P.e.a. Fahay. Allen Press, Lawrence, Kansas. pp. 142-149.

Kimura, M. 1980. A simple method for estimating evolutionary rate of base substitution through comparative studies of nucleotide sequences. J. Mol. Evol. **16**: 111-120.

Kinlan, B.P., and Gaines, S.D. 2003. Propagule dispersal in marine and terrestrial environments: A community perspective. Ecology **84**(8): 2007-2020.

Kinnison, M.T., and Hendry, A.P. 2004. From macro- to micro-evolution: tempo and mode in salmonid evolution. *In* Evolution illuminated: salmon and their relatives. *Edited by* A.P. Hendry and S.C. Stearns. Oxford University Press, Oxford, UK. pp. 208-231.

Koop, B., von Schalburg, K., Leong, J., Walker, N., Lieph, R., Cooper, G., Robb, A., Beetz-Sargent, M., Holt, R., Moore, R., Brahmbhatt, S., Rosner, J., Rexroad, C., McGowan, C., and Davidson, W. 2008. A salmonid E_{ST} genomic study: genes, duplications, phylogeny and microarrays. BMC Genomics **9**(1): 545.

Kornberg, A., Bertsch, L.L., Jackson, J.F., and Khorana, H.G. 1964. Enzymatic synthesis of deoxyribonucleic acid: XVI. Oligonucleotides as templates and the mechanisms of their replication. Proc. Natl. Acad. Sci. U.S.A. **51**: 315-323.

Kozak, K.H., Graham, C.H., and Wiens, J.J. 2008. Integrating GIS-based environmental data into evolutionary biology. Trends Ecol. Evol. **23**(3): 141-148.

Lagercrantz, U., Ellegren, H., and Andersson, L. 1993. The abundance of various polymorphic microsatellite motifs differs between plants and vertebrates. Nucleic Acids Res. **21**(5): 1111-1115.

Lemaire, C., Versini, J.J., and Bonhomme, F. 2005. Maintenance of genetic differentiation across a transition zone in the sea: discordance between nuclear and cytoplasmic markers. J. Evolution. Biol. **18**(1): 70-80.

Leonard, J.A., Wayne, R.K., and Cooper, A. 2000. Population genetics of Ice age brown bears. Proc. Natl. Acad. Sci. U.S.A. **97**(4): 1651-1654.

Lévêque, C., Oberdorff, T., Paugy, D., Stiassny, M.L.J., and Tedesco, P.A. 2008. Global diversity of fish (Pisces) in freshwater. Hydrobiologia **595**: 545-567.

Levins, R. 1969. Some demographic and genetic consequences of environmental heterogeneity for biological control. Bull. Entomol. Soc. Am. **15**: 237-240.

Levinton, J.S. 1979. A theory of diversity equilibrium and morphological evolution. Science **204**: 335-336.

Li, S., Pearl, D., and Doss, H. 2000. Phylogenetic tree reconstruction using Markov chain Monte Carlo. J. Am. Stat. Assoc. **95**: 493-508.

Li, W.H. 1976. Effect of migration on genetic distance. Am. Nat. **110**: 841-847.

Li, W.H. 1997. Molecular Evolution. Sinauer, Sunderland, MA. 487pp.

Lindsey, C.C. 1981. Stocks are chameleons: plasticity in gill rakers of coregonid fishes. Can. J. Fish. Aquat. Sci. **38**(12): 1497-1506.

Loehr, J., Worley, K., Grapputo, A., Carey, J., Veitch, A., and Coltman, D.W. 2006. Evidence for cryptic glacial refugia from North American mountain sheep mitochondrial DNA. J. Evolution. Biol. **19**(2): 419-430.

Lomolino, M.V., Sax, D.F., Riddle, B.R., and Brown, J.H. 2006. The island rule and a research agenda for studying ecogeographical patterns. J. Biogeogr. **33**(9): 1503-1510.

Lunt, D.H., and Hyman, B.C. 1997. Animal mitochondrial DNA recombination. Nature **387**(6630): 247-247.

MacArthur, R.H., Recher, H., and Cody, M. 1966. On the relation between habitat selection and species diversity. Am. Nat. **100**: 319-332.

MacArthur, R.H., and Wilson, E.O. 1963. An equilibrium theory of insular biogeography. Evolution **17**: 373-387.

Macpherson, A.H. 1965. The origin of diversity in mammals of the Canadian Arctic tundra. Syst. Zool.(14): 153-173.

Mann, G.J. 1974. Life history types of the least cisco (*Coregonus sardinella* Valenciennes) in Beaufort Sea drainages in the Yukon Territory, North Slope, and east Mackenzie River drainages. M.S. Thesis, University of Alberta, Edmonton, Alberta, Canada. 157pp.

Mann, G.J., and McCart, P.J. 1981. Comparison of sympatric dwarf and normal-populations of least cisco (*Coregonus sardinella*) inhabiting Trout Lake, Yukon Territory. Can J Fish Aquat Sci **38**(2): 240-244.

Marra, M.A., Jones, S.J.M., Astell, C.R., Holt, R.A., Brooks-Wilson, A., Butterfield, Y.S.N., Khattra, J., Asano, J.K., Barber, S.A., Chan, S.Y., Cloutier, A., Coughlin, S.M., Freeman, D., Girn, N., Griffith, O.L., Leach, S.R., Mayo, M., McDonald, H., Montgomery, S.B., Pandoh, P.K., Petrescu, A.S., Robertson, A.G., Schein, J.E., Siddiqui, A., Smailus, D.E., Stott, J.M., Yang, G.S., Plummer, F., Andonov, A., Artsob, H., Bastien, N., Bernard, K., Booth, T.F., Bowness, D., Czub, M., Drebot, M., Fernando, L., Flick, R., Garbutt, M., Gray, M., Grolla, A., Jones, S., Feldmann, H., Meyers, A., Kabani, A., Li, Y., Normand, S., Stroher, U., Tipples, G.A., Tyler, S., Vogrig, R., Ward, D., Watson, B., Brunham, R.C., Krajden, M., Petric, M., Skowronski, D.M., Upton, C., and Roper, R.L. 2003. The genome sequence of the SARS-associated coronavirus. Science **300**: 1399-1404.

Martin, P.D., Jenkins, J.L., Adams, F.J., Jorgenson, M.T., Matz, A.C., Payer, D.C., Reynolds, P.E., Tidwell, A.C., and Zelenak, J.R. 2009. Wildlife Response to Environmental Arctic Change: Predicting Future Habitats of Arctic Alaska. Report of the Wildlife Response to Environmental Arctic Change (WildREACH): Predicting Future Habitats of Arctic Alaska Workshop, 17-18 November 2008. Fish and Wildlife Service, Fairbanks, Alaska: U.S. 138pp.

Mäser, P., Thomine, S., Schroeder, J.I., Ward, J.M., Hirschi, K., Sze, H., Talke, I.N., Amtmann, A., Maathuis, F.J., Sanders, D., Harper, J.F., Tchieu, J., Gribskov, M., Persans, M.W., Salt, D.E., Kim, S.A., and Guerinot, M.L. 2001. Phylogenetic relationships within cation transporter families of Arabidopsis. Plant Physiol. **126**: 1646-1667.

Mau, B., and Newton, M.A. 1997. Phylogenetic inference for binary data on dendrograms using Markov chain Monte Carlo. J. Comp. Graph. Stat. **6**: 122-131.

McAlpine, S., Rhodes, O.E., McCreedy, C.D., and Brisbin, I.L. 1994. Genetic-structure in a wintering population of American coots. Wilson Bull. **106**(4): 738-743.

McMillan, W.O., and Palumbi, S.R. 1997. Rapid rate of control-region evolution in Pacific butterflyfishes (*Chaetodontidae*). J. Mol. Evol. **45**(5): 473-484.

McPhail, J.D., and Lindsey, C.C. 1970. Freshwater fishes of northwestern Canada and Alaska. Fisheries Research Board of Canada, Ottawa, Canada. 381pp.

Mecklenburg, C.W., Mecklenburg, T.A. and Thorsteinson, L.K. 2002. Fishes of Alaska. American Fisheries Society, Bethesda, MD, U.S.A.

Morrow, J.E. 1980. The Freshwater Fishes of Alaska. Alaska Northwest Publishing Company, Anchorage, AK. 248pp.

Moulton, L.L., Seavey, B., and Pausanna, J. 2010. History of an under-ice subsistence fishery for Arctic cisco and least cisco in the Colville River, Alaska. Arctic **63**(4): 381-390.

Muller, R.A., and MacDonald, G.J. 1997. Glacial cycles and astronomical forcing. Science **277**: 215-218.

Murphy, S.M., Mueter, F.J., and Braund, S.R. 2007. Variation in the abundance of Arctic cisco in the Colville River: Analysis of existing data and local knowledge. Technical Report No. MMS 2007-042. U.S. Department of the Interior, Minerals Management Service Alaska OCS Region, Anchorage, AK, U.S.A. 260pp.

Nedbal, M.A., and Flynn, J.J. 1998. Do the combined effects of the asymmetric process of replication and DNA damage from oxygen radicals produce a mutation-rate signature in the mitochondrial genome? Mol. Biol. Evol. **15**(2): 219-223.

Nielsen, E.E., Nielsen, P.H., Meldrup, D., and Hansen, M.M. 2004. Genetic population structure of turbot (*Scophthalmus maximus* L.) supports the presence of multiple hybrid zones for marine fishes in the transition zone between the Baltic Sea and the North Sea. Mol. Ecol. **13**(3): 585-595.

Norden, C.R. 1961. Comparative osteology of representative salmonid fishes, with particular reference to the grayling (*Thymallus arcticus*) and its phylogeny. J. Fish. Res. Board Can. **18**: 679-791.

O'Reilly, P.T., Canino, M.F., Bailey, K.M., and Bentzen, P. 2004. Inverse relationship between F-ST and microsatellite polymorphism in the marine fish, walleye pollock (*Theragra chalcogramma*): implications for resolving weak population structure. Mol. Ecol. **13**(7): 1799-1814.

Olsen, J.B., Crane, P.A., Flannery, B.G., Dunmall, K., Templin, W.D., and Wenburg, J.K. 2011. Comparative landscape genetic analysis of three Pacific salmon species from subarctic North America. Conserv. Genet. **12**(1): 223-241.

Østbye, K., Amundsen, P.A., Bernatchez, L., Klemetsen, A., Knudsen, R., Kristoffersen, R., Naesje, T.F., and Hindar, K. 2006. Parallel evolution of ecomorphological traits in the European whitefish *Coregonus lavaretus* (L.) species complex during postglacial times. Mol. Ecol. **15**(13): 3983-4001.

Phillips, R.B. and Rab, P. 2001. Chromosome evolution in the Salmonidae (Pisces): an update. Biol. Rev. **76**: 1-25.

Pielou, E.C. 1991. After the Ice Age: the Return of Life to Glaciated North America. University of Chicago Press, Chicago, IL. 376pp.

Prowse, T.D., Wrona, F.J., Reist, J.D., Gibson, J.J., Hobbie, J.E., Lévesque, L.M.J., and Vincent, W.F. 2006a. Climate change effects on hydroecology of Arctic freshwater ecosystems. Ambio **35**(7): 347-358.

Prowse, T.D., Wrona, F.J., Reist, J.D., Gibson, J.J., Hobbie, J.E., Lévesque, L.M.J., and Vincent, W.F. 2006b. Historical changes in Arctic freshwater ecosystems. Ambio **35**(7): 339-346.

Prowse, T.D., Wrona, F.J., Reist, J.D., Hobbie, J.E., Lévesque, L.M.J., and Vincent, W.F. 2006c. General features of the Arctic relevant to climate change in freshwater ecosystems. Ambio **35**(7): 330-338.

Rannala, B., and Yang, Z. 1996. Probability distribution of molecular evolutionary trees: a new method of phylogenetic inference. J. Mol. Evol. **43**: 304-311.

Reist, J.D., and Bond, W.A. 1988. Life history characteristics of migratory coregonids of the lower Mackenzie River, Northwest Territories, Canada. Finn. Fish. Res. **9**: 133-144.

Reist, J.D., Wrona, F.J., Prowse, T.D., Power, M., Dempson, J.B., Beamish, R.J., King, J.R., Carmichael, T.J., and Sawatzky, C.D. 2006a. General effects of climate change on Arctic fishes and fish populations. Ambio **35**(7): 370-380.

Reist, J.D., Wrona, F.J., Prowse, T.D., Power, M., Dempson, J.B., King, J.R., and Beamish, R.J. 2006b. An overview of effects of climate change on selected Arctic freshwater and anadromous fishes. Ambio **35**(7): 381-387.

Reshetnikov, J.S. 1988. Coregonid fishes in recent conditions. Finn. Fish. Res. **9**: 11-16.

Richter, C. 1992. Reactive oxygen and DNA damage in mitochondria. Mutat. Res. **275**(3-6): 249-255.

Ricklefs, R.E. 2004. A comprehensive framework for global patterns in biodiversity. Ecol. Lett. **7**(1): 1-15.

Ricklefs, R.E. 2006a. Evolutionary diversification and the origin of the diversity-environment relationship. Ecology **87**(7): S3-S13.

Ricklefs, R.E. 2006b. Global variation in the diversification rate of passerine birds Ecology **87**(10): 2468-2478.

Riddle, B.R. 1996. The molecular phylogeographic bridge between deep and shallow history in continental biotas. Trends Ecol. Evol. **11**(5): 207-211.

Rohde, K. 1992. Latitudinal gradients in species diversity - the search for the primary cause. Oikos **65**(3): 514-527.

Ross, C.L., Dyer, K.A., Erez, T., Miller, S.J., Jaenike, J., and Markow, T.A. 2003. Rapid divergence of microsatellite abundance among species of *Drosophila.* Mol. Biol. Evol. **20**(7): 1143-1157.

Rousset, F. 2002. Inbreeding and relatedness coefficients: what do they measure? Heredity **88**: 371-380.

Ruzzante, D.E., Taggart, C.T., and Cook, D. 1999. A review of the evidence for genetic structure of cod (*Gadus morhua*) populations in the NW Atlantic and population affinities of larval cod off Newfoundland and the Gulf of St. Lawrence. Fish. Res. **43**(1-3): 79-97.

Saitou, N., and Nei, M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4**: 406-425.

Salipante, S.J., and Horwitz, M.S. 2006. Phylogenetic fate mapping. Proc. Natl. Acad. Sci. U.S.A. **103**: 5448-5453.

Scott, W.B., and Crossman, E.J. 1973. Freshwater fishes of Canada. Fish. Res. Board Can. Bull. **184**: 966pp.

Seigle, J.C., and Parrett, J.P. 2008. Fall 2008 subsistence fishery monitoring on the Colville River. Report prepared for ConocoPhillips Alaska, Inc. by ABR, Inc., Anchorage, AK, U.S.A. 46pp.

Selkoe, K.A., and Toonen, R.J. 2006. Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. Ecol. Lett. **9**(5): 615-629.

Shafer, A.B.A., Cullingham, C.I., Cote, S.D., and Coltman, D.W. 2010. Of glaciers and refugia: a decade of study sheds new light on the phylogeography of northwestern North America. Mol. Ecol. **19**(21): 4589-4621.

Simon, J., Krauthoefer, T., Koster, D., and Caylor, D. 2007. Subsistence salmon harvest monitoring report, Kuskokwim Fisheries Management Area, Alaska, 2004. Alaska Department of Fish and Game, Division of Subsistence Technical Paper No. 313. Juneau, AK, U.S.A. 74pp.

Slatkin, M. 1991. Inbreeding coefficients and coalescence times. Genet. Res. **58**:167-175.

Slatkin, M. 1995. A measure of population subdivision based on microsatellite allele frequencies. Genetics **139**: 457-462.

Spitze, K. 1993. Population-structure in *Daphnia obtusa* - Quantitative genetic and allozymic variation. Genetics **135**(2): 367-374.

Stamford, M.D., and Taylor, E.B. 2004. Phylogeographical lineages of Arctic grayling (*Thymallus arcticus*) in North America: divergence, origins and affinities with Eurasian *Thymallus*. Mol. Ecol. **13**(6): 1533-1549.

Stearley, R.F., and Smith, G.R. 1993. Phylogeny of the Pacific trouts and salmon (*Oncorhynchus*) and genera of the family Salmonidae. Trans. Am. Fish. Soc. **122**: 1-36.

Steppan, S.J., Akhverdyan, M.R., Lyapunova, E.A., Fraser, D.G., Vorontsov, N.N., Hoffmann, R.S., and Braun, M.J. 1999. Molecular phylogeny of the marmots (Rodentia : Sciuridae): Tests of evolutionary and biogeographic hypotheses. Syst. Biol. **48**(4): 715-734.

Sunnucks, P. 2000. Efficient genetic markers for population biology. Trends Ecol. Evol. **15**(5): 199-203.

Sutton, T.M., and Edenfield, L.E. 2012. Stock characteristics of humpback whitefish and least cisco in the Chatanika River, Alaska. Arctic **65**(1): 67-75.

Swenson, N.G., and Howard, D.J. 2005. Clustering of contact zones, hybrid zones, and phylogeographic breaks in North America. Am. Nat. **166**(5): 581-591.

Takahata, N., and Nei, M. 1984. F_{st} and G_{st} statistics in the finite island model. Genetics **107**(3): 501-504.

Taylor, E.B. 2004. Evolution in mixed company: evolutionary inferences from studies of natural hybridization in Salmonidae. *In* Evolution illuminated: salmon and their relatives. *Edited by* A.P. Hendry and S.C. Stearns. Oxford University Press, Oxford, UK. pp. 232-263.

Thompson, S.L., and Whitton, J. 2006. Patterns of recurrent evolution and geographic parthenogenesis within apomictic polyploid Easter daises (*Townsendia hookeri*). Mol. Ecol. **15**(11): 3389-3400.

Tonn, W.M. 1990. Climate change and fish communities - a conceptual framework. Trans. Am. Fish. Soc. **119**(2): 337-352.

Tremblay, N.O., and Schoen, D.J. 1999. Molecular phylogeography of *Dryas integrifolia*: glacial refugia and postglacial recolonization. Mol. Ecol. **8**(7): 1187-1198.

Turgeon, J., and Bernatchez, L. 2003. Reticulate evolution and phenotypic diversity in North American ciscoes, *Coregonus* ssp (Teleostei : Salmonidae): implications for the conservation of an evolutionary legacy. Conserv. Genet. **4**(1): 67-81.

Van Oosterhout, C., Hutchinson, W.F., Wills, D.P.M., and Shipley, P. 2004. MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. Mol. Ecol. Notes **4**(3): 535-538.

Wallace, A.R. 1876. The Geographical Distribution of Animals: With a Study of the Relations of Living and Extinct Faunas as Elucidating the Past Changes of the Earth's Surface. Macmillian and Co., London. 574pp.

Wang, J. 2012. On the measurements of genetic differentiation among populations. Genet. Res. **94**(5): 275-289.

Ward, R.H., Redd, A., Valencia, D., Frazier, B., and Paabo, S. 1993. Genetic and linguistic differentiation in the America. Proc. Natl. Acad. Sci. U.S.A. **90**(22): 10663-10667.

Watts, P.C., O'Leary, D., Cross, M.C., Coughlan, J., Dillane, E., Kay, S.M., Wylde, S., Stet, R., Nash, R.D.M., Hatfield, E.M.C., and Cross, T.F. 2008. Contrasting levels of genetic differentiation among putative neutral microsatellite loci in Atlantic herring *Clupea harengus* populations and the implications for assessing stock structure. Hydrobiologia **606**: 27-33.

Watts, P.C., and Thorpe, J.P. 2006. Influence of contrasting larval developmental types upon the population-genetic structure of cheilostome bryozoans. Mar. Biol. **149**(5): 1093-1101.

Weir, B.S., and Cockerham, C.C. 1984. Estimating F-Statistics for the analysis of population-structure. Evolution **38**(6): 1358-1370.

Wilder, J., and Hollocher, H. 2001. Mobile elements and the genesis of microsatellites in dipterans. Mol. Biol. Evol. **18**(3): 384-392.

Wilson, A.C., Cann, R.L., Carr, S.M., George, M., Gyllensten, U.B., Helmbychowski, K.M., Higuchi, R.G., Palumbi, S.R., Prager, E.M., Sage, R.D., and Stoneking, M. 1985. Mitochondrial-DNA and 2 perspectives on evolutionary genetics. Biol. J. Linn. Soc. **26**(4): 375-400.

Wilson, C., and Hebert, P. 1998. Phylogeography and postglacial dispersal of lake trout (*Salvelinus namaycush*) in North America. Can. J. Fish. Aquat. Sci. **55**(4): 1010-1024.

Wilson, M.V.H., and Li, G.Q. 1999. Osteology and systematic position of the Eocene salmonid *Eosalmo driftwoodensis Wilson* from western North America. Zool. J. Linn. Soc.-Lond. **125**(3): 279-311.

Witt, J.D.S., Zemlak, R.J., and Taylor, E.B. 2011. Phylogeography and the origins of range disjunctions in a north temperate fish, the pygmy whitefish (*Prosopium coulterii*), inferred from mitochondrial and nuclear DNA sequence analysis. J. Biogeogr. **38**(8): 1557-1569.

Wright, S. 1943. Isolation by distance. Genetics **28**:114-138.

Wright, S. 1951. The genetical structure of populations. Ann. Eugen. **15**:323-354.

Wright, S. 1965. The interpretation of population structure by F-statistics with special regard to systems of mating. Evolution **19**:395-420.

Yang, Z., and Rannala, B. 1997. Bayesian phylogenetic inference using DNA sequences: a Markov chain Monte Carlo Method. Mol. Biol. Evol. **14**: 717-724.

Yang, Z., and Rannala, B. 2012. Molecular phylogenetics: principles and practice. Nat. Rev. Genet. **13**(5): 303-314.

Fig. 1.1. A least cisco (*Coregonus sardinella*) caught on the Arctic Coastal Plain (scale of photo: 4 = 400mm).

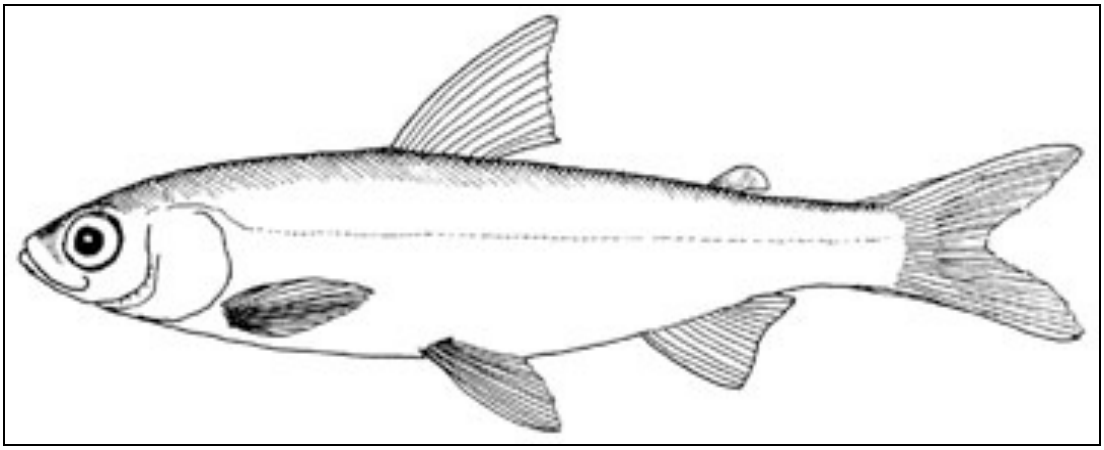

Fig. 1.2. Diagrammatic representation of a least cisco. From http://www.juvenilefishid.com/qm.htm (accessed 30 September 2012)**.**

Table 1.1. Interpretations of F_{ST}.

Equation	Interpretation	Reference
$F_{ST} = \frac{r_S - r_T}{1 - r_T}$	Inbreeding coefficient; a population parameter dependent on a population's demographic history; r_S and r_T are the correlations between two homologous genes drawn at random from within a subpopulation and from the total population, respectively; also known as the probability of identity by descent (PIBD)	(Wright 1943; Wright 1951; Wright 1965)
$F_{ST} = \frac{\sigma_p^2}{\bar{p}(1-\bar{p})}$	1. The proportion of genetic variation found between subpopulations at a locus; $\bar{p}$ and σ_p^2 are the mean and variance of allele frequencies in subpopulations, respectively	(Wright 1943; Wright 1965)
	2. A probability that two randomly chosen homologous genes both descended from a gene in the subpopulation	(Crow and Kimura 1970)
	3. An average level of inbreeding in the subpopulation relative to the whole population	(Falconer and Mackay 1996)
$F_{ST} = \frac{t_T - t_S}{t_T}$	Interpreted in terms of recent evolutionary history, with higher F_{ST} values indicating an increase in coalescence time between genes in the total population in relation to genes from a subpopulation; t_T and t_S are average coalescence times of two gene copies randomly drawn from the same subpopulation (t_S) and total population (t_T)	(Slatkin 1991; Slatkin 1995)

Chapter 2: Phylogeography of least cisco (*Coregonus sardinella*) in Alaska[1]

Abstract

In this study, a combination of analyses helped build a first detailed phylogeographic analysis of least cisco (*Coregonus sardinella*) in Alaska, which suggests that the least cisco's evolutionary history differs from other *Coregonus* species in Alaska. The least cisco is relatively diverse across Alaska, 68 unique haplotypes in 305 individuals, and Φ_{ST} values indicate incipient genetic differentiation. The haplotype network and phylogeny show little evidence of geographic segregation among haplotypes, suggesting high levels of historic gene flow or recent colonization. But some haplotype groups in the network are restricted to a region, and it is not possible to draw a one to one match between haplotype groups and regions when examining the entire network. Overall, the data show that a large proportion of genetic variation is shared across Alaska, but this variation is not homogenously distributed across all regions and for all haplotype groups.

Introduction

Least cisco (*Coregonus sardinella*)

Least cisco, salmonids in the group commonly known as the coregonids, or whitefishes, are distributed across Arctic Canada, Alaska, and Siberia. In Alaska, their freshwater habitats range from the Arctic Coast to Bristol Bay. They inhabit the rivers, streams, and lakes north of the Alaska Range that drain into the Arctic Ocean, and the rivers, streams, and lakes south of the Alaska Range that drain into the Bering Sea. This broad Alaskan distribution, combined with diverse migration patterns, means that least cisco encounter highly variable environmental conditions. Least cisco exhibit high levels of phenotypic and life history plasticity (Harper et al. 2007), likely equipping them for survival in the extreme environmental conditions of the Alaskan hydrographic landscape.

---

[1] Padula, V. M., D. Causey, and J. A. Lopez. 2013. Chapter 2: Phylogeography of least cisco (*Coregonus sardinella*) in Alaska. Prepared for submission to the Canadian Journal of Fisheries and Aquatic Sciences.

Many evolutionary relationships remain unresolved within the Coregoninae. Currently, 28 species are recognized in the genus *Coregonus*, the highest number of any salmonid genus (Reshetnikov 1988). New and Old World coregonid populations aggregate into six distinct genetic lineages: (i) Inconnu (genus *Stenodus*); (ii) an Arctic cisco group; (iii) a European (*C. lavaretus*) and lake whitefish (*C. clupeaformis*) group; (iv) broad whitefish (*C. nasus*); (v) a least cisco and vendace (*C. albula*) group; and (vi) a peled (*C. peled*) group (Bodaly et al. 1991). Relatively little information exists about the placement of least cisco within the broader Coregoninae in particular. Furthermore, the genetic relationships among Alaskan populations within the species are almost entirely unknown. Lack of resolution results from difficulty in species identification (Crête-Lafrenière et al. 2012), parallel evolution (Bernatchez 2004; Østbye et al. 2006), phenotypic plasticity (Lindsey 1981), recurrent trophic polymorphisms (Bernatchez et al. 1991; Douglas et al. 1999), contemporary hybridization (Taylor 2004), and historical introgression (Turgeon and Bernatchez 2003). To highlight this point, the widely accepted morphological groupings of coregonids into whitefish and ciscoes may also be inaccurate (Bernatchez et al. 1991), as recent analyses have shown that members of the sardine cisco clade (*C. sardinella, C. albula* and *C. peled*) are grouped with 'true' whitefish species (Crête-Lafrenière et al. 2012).

Alaskan freshwater biogeography

Taxonomic confusion stems from phenotypic plasticity and diversity within and among species, instances of species flocks and sibling species with restricted distributions, and unusual distributional patterns often related to glaciations within their ranges (Bodaly et al. 1992; Lindsey 1981). The influence of glacial advances and retreats is a particularly important factor when considering the underlying processes shaping the phylogeographic patterns we see among coregonids today because coregonids inhabit northern aquatic regions that were greatly affected by the Pleistocene ice age. Much of Canada was covered by ice-sheets during the Pleistocene, with the largest ice-sheet coverage occurring approximately 20,000 years ago (Clark et al. 2009); the two major ice refugia were Beringia and the Pacific Northwest (Shafer et al. 2010). The glacial cycles and

climatic fluctuations of the Pleistocene are one of the underlying mechanisms shaping the complex phylogeographic patterns of northwestern North American biota (Shafer et al. 2010).

Phylogeographic studies show that Beringia, which encompassed Alaska, was a refuge for numerous species, particularly freshwater species. For example, a freshwater refugium of periglacial freshwater lakes likely existed for lake trout, *Salvelinus namaycush* (Wilson and Hebert 1998), and Arctic grayling, *Thymallus arcticus* (Stamford and Taylor 2004), in the southeastern portion of Beringia at the juncture of the Cordilleran and Laurentide ice-sheets (Dyke and Prest 1987). Breaks in ice-sheets and the Yukon River are proposed barriers that created phylogeographic breaks (Eddingsaas et al. 2004; Jorgenson et al. 2003). Populations of northwestern North America's biota survived in glacial refugia and other suitable habitats (Holderegger and Thiel-Egenter 2009) by shifting their ranges across refugia. These groups were often isolated and genetically differentiated from conspecific populations over time, maintaining genetic diversity (Shafer et al. 2010). As glaciers retreated, these groups expanded into newly available terrain, often encountering and mixing with other conspecifics (Hewitt 2001). This process may has been repeated several times during the many glacial advances and retreats that occurred during the last 700,000 years.

Study goals

Least cisco populations have been studied in the various major river drainages along the Arctic Alaskan and Canadian coasts, but little or no previous genetic research has been reported for this species. The information presented here provides an initial baseline of genetic relationships among least cisco populations across Alaska. As climate change alters Alaskan freshwater habitats and change connections and barriers to gene flow for freshwater fishes, future assessments of genetic relationships among populations can be compared to this baseline information.

We addressed the question of how least cisco populations in different regions of Alaska are genetically related each other. Therefore, the objectives of this research were to: (i) characterize mtDNA variability in nine Alaskan populations of least cisco,

including the Arctic Coastal Plain (ACP), Point Lay (LAY), Barrow (BAR), Selawik River (SEL), Koyukuk River (KOY), Rampart Rapids (RAP), Lake Clark (CLA), Whitefish Lake (WHI), and Toksook River (TOK); (ii) determine the degree of differentiation among populations; and (iii) if differentiation is present, estimate the timing of separation and/or suggest barriers to movement.

Materials and methods

Sample collection

Least cisco were sampled from nine locations: the Arctic Coastal Plain (ACP, *n* = 163), Point Lay (LAY, *n* = 12), Barrow (BAR, *n* = 27), and Selawik River (SEL, *n* = 29). Locations in the Bering Sea Drainage included: Koyukuk River (KOY, *n* = 13), Rampart Rapids (RAP, *n* = 19), Lake Clark (CLA, *n* = 15), Whitefish Lake (WHI, *n* = 22), and Toksook River (TOK, *n* = 5) (Fig. 2.1, Table 2.2). Fish were captured using fyke net, gill net, minnow trap, fish wheel, or surface trawl. A pelvic fin was clipped from each fish and the tissue samples were stored individually in vials and preserved in 95% ethanol and stored at -20°C until processing.

DNA sequencing

DNA was isolated from the Arctic Coastal Plain and Point Lay samples using the Puregene DNA Extraction Core Kit (Qiagen, Germantown, MD) following the manufacturer's protocol. DNA was extracted from the Barrow, Selawik River, Koyukuk River, Rampart Rapids, Lake Clark, Whitefish Lake, and Toksook River samples using a DNeasy Blood and Tissue Kit (Qiagen, Germantown, MD) following the manufacturer's protocol. The salmonid control region (D-loop) of the mitochondrial genome was isolated and amplified by polymerase chain reaction (PCR). Conditions for a 25 μL reaction volume were as follows: 0.5-1 μL DNA template (concentration of approximately 100 ng/mL), 1X GoTaq^{TM} (Promega Corporation, Madison, WI) buffer, 0.8 mM dNTPs, 2.5 mM $MgCl_2$, 1X BSA, 0.2 μM forward and reverse primer, and 0.025 U/μL GoTaq (Promega Corporation, Madison, WI). Reactions were run for 32 cycles with an annealing temperature of 55°C, a 30 s extension at 72°C and a 10 m post-incubation at

72°C. PCR products were checked by gel electrophoresis to verify amplifications of the correct product on a 2.5% agarose gel.

Individuals were sequenced on an automated DNA sequencer (ABI 3100, University of Washington DNA). Chromatograms were edited in CodonCode Aligner 4.1 (CodonCode Corporation, Centerville, MA) to remove noise and artifacts. Forward and reverse sequence reads were assembled in Mesquite (Maddison and Maddison 2010), and all samples were aligned in MEGA (Tamura 2011) using a ClustalW algorithm (Thompson et al. 1994) with the default settings of 15 gap opening penalty, 6.66 gap extension penalty for both pairwise and multiple alignment, 0.5 transition weight, and 30% delay divergent cutoff. Alignment allowed for identification and verification of indels in a subsample of sequences. Gaps greater than one bp appeared in two locations. At bp 623-624 (position determined by counting the first base in the sequences as one), sequences either had TA or a two bp gap. Site 623 was therefore recoded to T or G, respectively. At bp 640-657, samples had one of three variants: (i) a sequence of ATTAATAAAATTTATTGC; (ii) a sequence of ATTAATAAAACTTATTGC; or (3) a gap of 18 bp. Site 639 was recoded to A, C, or G, respectively.

Phylogeographic analyses

The resulting sequence data were analyzed in three ways. First, the variability and divergence were characterized for the sequence data. Second, haplotype networks were constructed and Φ_{ST} (an analogue to F_{ST}) values were estimated for between and within drainage comparisons. Finally, a phylogeny of haplotypes with bootstrap values of support was constructed. For the first set of analyses, unique haplotypes were identified for the entire sample set, for each region, and for each location within those regions in DnaSP (Librado and Rozas 2009). Haplotype diversity (H_d) and the average number of differences (K) were also estimated for the entire sample set, for each region, and for each location within those regions in DnaSP (Librado and Rozas 2009). Values of evolutionary divergence between sequences was measured as base substitutions per site between sequences. This was done in MEGA using a maximum composite likelihood method (Tamura et al. 2004), where rates of substitution were assumed to be uniform

among sites (Jukes-Cantor model). The same was done between all pairs of unique haplotypes. The frequencies of evolutionary divergence values were calculated in Excel (Microsoft Corporation, Redmond, WA) and graphed. Haplotypic richness values were calculated in FSTAT (Goudet 1995) for the nine sampling sites.

A minimum spanning network of haplotypes was calculated using TCS (Clement et al. 2000), where all sequences were collapsed into haplotypes, an absolute distance matrix was calculated for all pairwise comparisons of haplotypes, and then a parsimony connection limit was calculated, resulting in a set of justified connections. A haplotype network image was created in Network ver. 4.610 (fluxus-engineering.com) using the minimum spanning network data.

Pseudoexact *G* tests for homogeneity were then performed in ARLEQUIN version 3.5.1.3 (Excoffier and Lischer 2010) to determine if differences in genetic compositions existed across the nine sampling sites. Between-population Φ_{ST} (an analogue to F_{ST}) values were estimated in ARLEQUIN version 3.5.1.3 (Excoffier and Lischer 2010), with a significance level of 0.05, 100 permutations for significance, 10,100 permutations for Mantel test. Exact tests of population differentiation were estimated with 100,000 MCMC steps and 10,000 dememorization steps, with sequential Bonferroni corrections applied to *p*-values to account for multiple testing.

A maximum likelihood (ML) parsimony phylogeny reconstruction was performed in MEGA, by using a bootstrap method of 10,000 replications, a general time reversible (GTR) model of nucleotide substitution, uniform rates among sites (Jukes-Cantor model), and nearest-neighbor-interchange (NNI) heuristic method. The following sequence accessions were downloaded from GenBank and used as outgroups: JQ661394.1 *C. lavaretus* Ro4, JQ661396.1 *C. lavaretus* Ro6, JQ661388.1 *C. lavaretus* Est7, JQ661382.1 *C. lavaretus* Est6, JQ661384.1 *C. lavaretus* Est10, HM634595.1 *C. albula* Alb15, HM634594.1 *C. albula* Alb13, HM634593.1 *C. albula* Alb09, HM634592.1 *C. albula* Alb05, HM634591.1 *C. albula* Alb03, HM635009.1 *C. renke* WT03A, HM635008.1 *C. renke* WT02A, HM635007.1 *C. renke* WT01A, HM634871.1 *C. maraena* SH11A, HM634692.1 *C. maraena* BSH771A, HM634761.1 *C. holsatus* GM29B, HM634699.1 *C.*

widegreni BSS04A, HM634695.1 *C. maraena* BSH775A, HM634650.1 *C. macrophthalmus* BS03A, HM635016.1 *C. renke* WES05A, HM634956.1 *C. fatioi* TR18A, HM634899.1 *C. suidteri* SEBA214, HM634850.1 *C. palaea* NC08A, HM635060.1 *C. duplex* ZU28B, HM635030.1 *C. heglingus* WASA102, HM634862.1 *C. candidus* NEBO302. Adding the outgroup sequences to the pre-existing alignment of *C. sardinella* sequences created new gaps at the following sites: 446-465, 473, 507, 562, 603, 604, 623-624, 623, 640-657. Gaps were recoded in the same way as before, resulting in a 631 bp sequence.

The lengths of specific branches were measured (in substitutions/site) on the resulting phylogeny and used to approximate times of divergence. Average mitochondrial rates of molecular divergence were recently estimated by Crête-Lafrenière et al. (2012), The average rate across genera in Salmonidae, based on an analysis of the Cytb gene and a segment of the CO1 gene, is approximately estimated at 0.31%/MY (CI: 0.27–0.36%/MY) (Crête-Lafrenière et al. 2012). However, the study found that the mitochondrial rate of molecular divergence varies markedly among genes and clades (Crête-Lafrenière et al. 2012), and the divergence rates for *Coregonus* ranged between 0.20%/MY (CI: 0.17–0.23%/MY) and 0.33%/MY (CI: 0.27–0.38%/MY) (Crête-Lafrenière et al. 2012). We used a midpoint of this range, 0.26%/MY, as the mitochondrial rate of molecular divergence to approximate time of divergence as indicated by the branches of the haplotype phylogeny. We acknowledge that these rates may not exactly reflect the rate of mutation for the mitochondrial control region, but this publication is the most recent study of salmonid mtDNA that was applicable to the present study.

Results

Sequence variability

We sequenced the mitochondrial control region (678 bp) of 305 *C. sardinella* individuals. We identified 68 unique haplotypes (Table 2.1), and the average evolutionary divergence was 0.011 ± 0.007% (substitutions/site), the overall haplotype diversity (H_d) was 0.935 with an average number of differences (K) was 8.266 (Table 2.3). Although

the most unique haplotypes were found in ACP, SEL had a higher haplotypic richness (12.566) than ACP (11.147) (Table 2.3). The maximum pairwise evolutionary divergence between all unique haplotypes and between all samples was 0.027%. The most frequent value was 0.019 when the frequencies of all the pairwise genetic distances were calculated (between all haplotypes and between all samples) (Fig. 2.2).

In ACP, 32 unique haplotypes were identified among 163 samples, with an average evolutionary divergence of 0.012 ± 0.008%, H_d = 0.894, and K = 8.379; 13 unique haplotypes were identified among 27 samples from BAR, with an average evolutionary divergence of 0.011 ± 0.009%, H_d = 0.795, and K = 8.188; 4 unique haplotypes were identified among 12 samples from LAY, with an average evolutionary divergence of 0.010 ± 0.008,% H_d = 0.758, and K = 7.606; 17 unique haplotypes were identified among 29 samples from SEL, with an average evolutionary divergence of 0.010 ± 0.006%, H_d = 0.916, and K = 7.337; 7 unique haplotypes were identified among 13 samples from KOY, with an average evolutionary divergence of 0.011 ± 0.007%, H_d = 0.872, and K = 7.487; 7 unique samples were identified among 19 samples from RAP, with an average evolutionary divergence of 0.009 ± 0.007%, H_d = 0.807, and K = 6.070; 5 unique haplotypes were identified among 15 samples from CLA, with an average evolutionary divergence of 0.003 ± 0.003%, H_d = 0.790, and K = 2.000; 3 unique haplotypes were identified among 5 samples from TOK, with an average evolutionary divergence of 0.009 ± 0.007%, H_d = 0.700, and K = 6.400; 13 unique haplotypes were identified among 22 samples from WHI, with an average evolutionary divergence of 0.009 ± 0.006%, H_d = 0.931, and K = 6.506 (Table 2.3).

Haplotype network and *Φ*-statistics

The largest genetic distance at the haplotype level was 0.027, calculated when haplotypes 49 and 53 were compared to haplotypes 2, 3, 15, 27, and 59. Haplotypes 49 and 53 radiated from haplotype 16 in the upper left corner of the network (Fig. 2.3), while the other haplotypes were located in the lower right corner of the figure, with haplotypes 2, 3 and 27 radiating from haplotype 4, the haplotype that occurred with the highest frequency in the entire dataset. The BAR group presented an interesting pattern in

the network. Of the 27 samples in BAR, 17 were collected at the freshwater site, and 10 were collected at Elson Lagoon. Within the freshwater group, five individuals had haplotypes 16, 49, or 50, which were in the upper left corner of the haplotype network (Fig. 2.3), but the other 12 individuals represented haplotype 4, located in the lower right corner of the network in the lineage most distantly related to the haplotype 16 lineage. Meanwhile, 10 haplotypes were found among the 10 individuals from Elson Lagoon, and these haplotypes were spread across the network, representing all the lineages found. Individuals representing haplotypes 49 and 27 revealed that both were from the Barrow population. The genetic distance between these two individuals was the highest calculated among pairwise comparisons, yet came from the same geographic region. On a finer scale, samples from the BAR group were collected at two sampling locations, one a freshwater site and the other Elson Lagoon, which has easier access to marine waters. Haplotype 49 was found in a fish collected at the freshwater site, while haplotype 27 was found in a fish collected at Elson Lagoon, begging a further investigation into the genetic variability of the fish representing the BAR group. The homogeneity test results reported in Table 2.4 suggest three groupings. ACP, BAR, and LAY had non-significant values when compared to each other. SEL, KOY, and RAP had non-significant values when compared to each other. CLA, WHI, and TOK had non-significant values when compared to each other. Likewise, those three groupings had low and non-significant Φ_{ST} values, reported in Table 2.5.

Phylogeny

The outgroup sequences (in green) were distinct from the *C. sardinella* sequences, with a bootstrap value of 0.7066 (Fig. 2.4). A polytomy appeared among the *C. sardinella* haplotypes, with haplotypes 1-59 forming the large group at the top of the phylogeny, and haplotypes 60-68 forming their own small clade in between that large group and the outgroups. The bootstrapping values separating the 1-59 group from the 60-68 group were 0.7533 and 0.5007 respectively, lending support for these groupings. This geographic variation coupled with low bootstrapping values may suggest a high degree of genetic mixing across regions, although the weak bootstrapping values were

probably more likely related to the fact that only 305 individuals were sequenced, which is a relatively low number in this type of study. Key sites could be missing from many of the iterations since the bootstrapping process samples data with replacement. The 60-68 haplotypes were all in blue, suggesting that this genetic clade may be unique to the far northern regions of Alaska.

We calibrated the time since divergence by measuring the branch length of the most divergent haplotype in Fig. 2.4. This was Hap_59, and the cumulative distance of this branch from the beginning point of the phylogeny was 0.0229 substitutions/site. Using 0.26%/MY as the mitochondrial rate of molecular divergence, this branch represented approximately 8.81MY since this group of least cisco diverged from the *Coregonus* species used in the outgroup. The other two branches of interest on the phylogeny were those leading to the 1-59 clade and the 60-68 clade. The length of the branch leading to the 1-59 clade was 0.0069 substitutions/site, representing approximately 6.15MY since divergence, and the length of the branch leading to the 60-68 clade was 0.0019, representing approximately 8.08MY since divergence.

Discussion

Least cisco are relatively diverse across Alaska, with 68 unique mitochondrial haplotypes found in 305 individuals. Homogeneity tests and Φ_{ST} values supported the idea that populations are isolated in the present day, and indicated incipient genetic differentiation across Alaska. No strong geographic segregation patterns could be detected in the 1-59 group of the haplotype network and phylogeny, suggesting some level of gene flow following an earlier isolation event. But some haplotype groups in the network were restricted to a region, and it was not possible to construct a one-to-one match between haplotype groups and regions when examining the entire network. Overall, the data showed that a large proportion of genetic variation was shared across Alaska, but this variation was not homogenously distributed across all regions and for all haplotype groups, and we may be observing the beginning of lineage sorting for Alaskan least cisco.

The maximum likelihood phylogeny showed a polytomy within *C. sardinella*. Haplotypes 60-68 form a distinct clade at the root of the tree, while haplotypes 1-59 form

another distinct clade. Haplotypes 60-68 all occurred in the northern region, while haplotypes 1-59 were a mixture across Alaska, with little evidence of geographic segregation among haplotypes. The distinction of the 60-68 group was further supported in the haplotype networks in Fig 2.3, which also showed they are not exclusive to one of the more northern sampling locations. Two hypotheses may explain this polytomy pattern: (i) the 60-68 clade evolved in sympatry with the 1-59 clade; or (ii) the 60-68 clade evolved in isolation from the 1-59 clade, perhaps in a glacial refugium, and the present-day pattern of polytomy is the result of secondary contact through migration and recolonization.

The high degree of genetic mixing among the 1-59 group suggested by the phylogeny and haplotype networks was not the entire picture of least cisco phylogeography in Alaska. The homogeneity tests and Φ-statistics revealed another detail about Alaskan least cisco. Although we did not originally group the nine populations into a broader hierarchical structure, results of homogeneity tests suggested that the nine populations used in this study could be treated as three regions. The most northern region consisted of ACP, LAY, and BAR. The middle region consisted of SEL, KOY, and RAP. The most southern region consisted of CLA, WHI, and TOK. This hierarchical structure was further supported by the low and non-significant Φ_{ST} values within those regions, and the higher and significant Φ_{ST} values between the regions (Table 2.5). The differentiation found among these populations may be the consequence of geographical barriers. For example, the Brooks Range and the Alaska Range are sizeable geographic barriers between the groups, and most likely limit gene flow.

Although Φ_{ST} should be an unbiased measure of genetic differentiation, factors such as the demographic history of a population (e.g. effective population size and migration rate), properties of the genetic markers (e.g. diversity and mutations) and sampling properties (e.g. sample size) can introduce bias (Wang 2012). The sample size for ACP was an order of magnitude larger than the sample sizes for any other subpopulation (163 sequences vs. sample sizes of 42 sequences or less for the other subpopulations). ACP received heavier weighting in Φ_{ST} calculations, even though the

sample sizes did not reflect what is known about the effective population sizes of these subpopulations, and were simply an artifact of what samples were available for analysis.

The phylogeography of least cisco in this study presented complex patterns that were likely shaped by glacial refugia in Beringia. Bernatchez and Dodson (1994) suggested similar hypotheses to explain the phylogeographic patterns they observed for lake whitefish (*C. clupeaformis*) mtDNA in Beringia. Lake whitefish mtDNA diversity was highest in Beringia (in comparison to lake whitefish from the rest of northern North America used in the study), with two distinct clades found within the region. The clades were mixed, showing no obvious geographic structure other than the differences in haplotype composition. Other whitefish species showed similar patterns. For example, previous studies of pygmy whitefish *P. coulterii* across their North American range suggested a complex history of post-glacial dispersal during the Wisconsin glaciations, with populations isolated in three refugia: a Beringian Refuge (Alaskan and possibly Yukon populations); a Pacific Refuge (Cascadian, Peace/Mackenzie and possibly Yukon populations); and a Mississippi Refuge (Lake Superior population) (Bailey and Smith 1981; Lindsey and Franzin 1972; Lindsey and McPhail 1986; McPhail and Lindsey 1986; Underhill 1986). A later study of mtDNA variation of pygmy whitefish from the same region suggested two clades, with one group derived from populations in Alaska and Aishihik Lake, and another group derived from Cascadia, the Peace River drainage, and Lake Superior (Witt et al. 2011). Other freshwater fishes that likely survived in the Beringian Refuge like the pygmy whitefish include lake whitefish *Coregonus clupeaformis* (Bernatchez and Dodson 1994), lake trout *Salvelinus namaycush* (Wilson and Hebert 1998), Arctic char *Salvelinus alpinus*, (Brunner et al. 2001), and Arctic grayling *Thymallus arcticus* (Stamford and Taylor 2004).

Finally, our estimates of divergence for least cisco (8.8MY) were similar to the estimated time of divergence for the *C. sardinella, C. peled,* and *C. albula* clade reported in Crête-Lafrenière et al. (2012) (e.g. ~8.5 MY; see Figure 6). Our fine-scale investigation of least cisco phylogenetics, using one locus from the mitochondrial genome, tracked well with this most recent analysis of the phylogenetic relationships

within the Salmonidae family resulting from using a number of genes, although we acknowledge difficulties arose in this comparison because of different mutation rates among genes. Adding nuclear genealogies to our analysis in the future could help draw stronger conclusions about the placement of least cisco within a broader phylogenetic context. Additionally, sampling genealogies from the nuclear genome may allow for the study of differential gene flow among loci because if locus-specific patterns of differentiation are mediated by selection, they can help explicate speciation mechanisms (Hare 2001). The emergence of next-generation sequencing technologies will enable multilocus phylogenetic analyses and large-scale comparative genomic studies that will provide more insight about the genome. For example, comparative mapping may reveal chromosomal rearrangements or colinearity that could not be detected otherwise (Cristescu et al. 2010).

Acknowledgements

Samples from ACP were collected in 2009 and 2010 from the freshwater lakes on the National Petroleum Reserve of Alaska (NPR-A) under the permit: [148907-1] The distribution and structure of fish communities in arctic lakes on the North Slope, Alaska. Ora Schlei at the USFWS Conservation Genetics Laboratory (Anchorage, AK) provided samples from SEL, KOY, RAP, WHI, and TOK, which were collected under State of Alaska Department of Fish and Game Fish Resource Permits for Scientific/Educational Purposes. Samples from BAR and LAY were collected in 2010 by Daniel Rizzolo at UAF under UAF IACUC assurance number 08-28 and ADFG Scientific Permit 10-006. Christian Zimmerman at USGS (Anchorage, AK) provided least cisco from CLA, which were caught under the following permits: Fish Resource Permit from Alaska Department of Fish and Game, Research Permit from National Park Service, and IACUC review by the USGS Alaska Science Center Animal Care and Use Committee.

References

Bailey, R.M., and Smith, G.R. 1981. Origin and geography of the fish fauna of the Laurentian Great Lakes Basin. Can. J. Fish. Aquat. Sci. **38**: 1539-1561.

Bernatchez, L. 2004. Ecological theory of adaptive radiation: an empirical assessment from coregonine fishes (Salmoniformes). *In* Evolution illuminated: salmon and their relatives. *Edited by* A.P. Hendry and S.C. Stearns. Oxford University Press, Oxford, UK. pp. 175-207.

Bernatchez, L., Colombani, F., and Dodson, J.J. 1991. Phylogenetic relationships among the subfamily Coregoninae as revealed by mitochondrial DNA restriction analysis. J. Fish Biol. **39**(A): 283-290.

Bernatchez, L., and Dodson, J.J. 1994. Phylogenetic relationships among palearctic and nearctic whitefish (*Coregonus* sp) populations as revealed by mitochondrial DNA variation. Can. J. Fish. Aquat. Sci. **51**(1): 240-251.

Bodaly, R.A., Clayton, J.W., Lindsey, C.C., and Vuorinen, J. 1992. Evolution of lake whitefish (*Coregonus clupeaformis*) in North America during the Pleistocene - genetic differentiation between sympatric populations. Can. J. Fish. Aquat. Sci. **49**(4): 769-779.

Bodaly, R.A., Vuorinen, J., Wards, R.D., Luczynski, M., and Reist, J.D. 1991. Genetic comparisons of New and Old-World coregonid fishes. J. Fish Biol. **38**(1): 37-51.

Brunner, P.C., Douglas, M.R., Osinov, A., Wilson, C.C., and Bernatchez, L. 2001. Holarctic phylogeography of Arctic charr (*Salvelinus alpinus L.*) inferred from mitochondrial DNA sequences. Evolution **55**(3): 573-586.

Clark, P.U., Dyke, A.S., Shakun, J.D., Carlson, A.E., Clark, J., Wohlfarth, B., Mitrovica, J.X., Hostetler, S.W., and McCabe, A.M. 2009. The last glacial maximum. Science **325**(5941): 710-714.

Clement, M., Posada, D., and Crandall, K.A. 2000. TCS: a computer program to estimate gene genealogies. Mol. Ecol. **9**(10): 1657-1660.

Crête-Lafrenière, A., Weir, L.K., and Bernatchez, L. 2012. Framing the salmonidae family phylogenetic portrait: A more complete picture from increased taxon sampling. Plos One **7**(10): 1-19.

Cristescu, M.E., Adamowicz, S.J., Vaillant, J.J., and Haffner, D.G. 2010. Ancient lakes revisited: from the ecology to the genetics of speciation. Mol. Ecol. **19**(22): 4837-4851.

Douglas, M., Brunner, P., and Bernatchez, L. 1999. Do assemblages of *Coregonus* (Teleostei : Salmoniformes) in the Central Alpine region of Europe represent species flocks? Mol. Ecol. **8**(4): 589-603.

Dyke, A.S., and Prest, V.K. 1987. The Late Wisconsinan and Holocene history of the Laurentide Ice Sheet. Geogr. Phys. Quatern. **41**: 237-263.

Eddingsaas, A.A., Jacobsen, B.K., Lessa, E.P., and Cook, J.A. 2004. Evolutionary history of the Arctic ground squirrel (*Spermophilus parryii*) in Nearctic Beringia. J. Mammal. **85**(4): 601-610.

Excoffier, L., and Lischer, H.E.L. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol. Ecol. Resour. **10**(3): 564-567.

Goudet, J. 1995. FSTAT (Version 1.2): A computer program to calculate *F*-statistics. J. Hered. **86**(6): 485-486

Hare, M.P. 2001. Prospects for nuclear gene phylogeography. Trends Ecol. Evol. **16**(12): 700-706.

Harper, K.C., Harris, F., Brown, R.J., Wyatt, T., and Cannon, D. 2007. Stock assessment of broad whitefish, humpback whitefish and least cisco in Whitefish Lake, Yukon Delta National Wildlife Refuge, Alaska, 2001-2003. Alaska Fisheries Technical Report Number 88. U.S. Fish and Wildlife Service, Fairbanks Fish and Wildlife Field Office. 55pp.

Hewitt, G.M. 2001. Speciation, hybrid zones and phylogeography - or seeing genes in space and time. Mol. Ecol. **10**(3): 537-549.

Holderegger, R., and Thiel-Egenter, C. 2009. A discussion of different types of glacial refugia used in mountain biogeography and phylogeography. J. Biogeogr. **36**(3): 476-480.

Jorgenson, M.T., Roth, J.E., Schlentner, S.F., and Cater, T.C. 2003. Ecological land evaluation for the Yukon Training Area on Fort Wainwright: permafrost and disturbance. Available from ABR, Fairbanks, Alaska, for U.S. Army, Anchorage, Alaska. 88pp.

Librado, P., and Rozas, J. 2009. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. Bioinformatics **25**: 1451-1452.

Lindsey, C.C. 1981. Stocks are chameleons: Plasticity in gill rakers of coregonid fishes. Can. J. Fish. Aquat. Sci. **38**(12): 1497-1506.

Lindsey, C.C., and Franzin, W.G. 1972. New complexities in the zoogeography and taxonomy of pygmy whitefish (*Prosopium coulteri*). J. Fish. Res. Board Can. **29**: 1772-1775.

Lindsey, C.C., and McPhail, D.J. 1986. Zoogeography of fishes of the Yukon and Mackenzie basins. *In* The zoogeography of North American freshwater fishes. *Edited by* C.H. Hocutt and E. O.Wiley. John Wiley & Sons, New York. pp. 639-674.

Maddison, W.P., and Maddison, D.R. 2010. Mesquite: a modular system for evolutionary analysis. http://mesquiteproject.org.

McPhail, J.D., and Lindsey, C.C. 1986. Zoogeography of the freshwater fishes of Cascadia (the Columbia system and rivers north to the Stikine). *In* The zoogeography of North American freshwater fishes. *Edited by* C.H. Hocutt and E. O.Wiley. John Wiley & Sons, New York. pp. 615-638.

Reshetnikov, J.S. 1988. Coregonid fishes in recent conditions. Finn. Fish. Res. **9**: 11-16.

Shafer, A.B.A., Cullingham, C.I., Cote, S.D., and Coltman, D.W. 2010. Of glaciers and refugia: a decade of study sheds new light on the phylogeography of northwestern North America. Mol. Ecol. **19**(21): 4589-4621.

Stamford, M.D., and Taylor, E.B. 2004. Phylogeographical lineages of Arctic grayling (*Thymallus arcticus*) in North America: divergence, origins and affinities with Eurasian Thymallus. Mol. Ecol. **13**(6): 1533-1549.

Tamura, K. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. **28**: 2731-2739.

Tamura K., Nei M., and Kumar S. (2004). Prospects for inferring very large phylogenies by using the neighbor-joining method. Proc. Natl. Acad. Sci. U.S.A. **101**:11030-11035.

Taylor, E.B. 2004. Evolution in mixed company: evolutionary inferences from studies of natural hybridization in Salmonidae. *In* Evolution illuminated: salmon and their relatives. *Edited by* A.P. Hendry and S.C. Stearns. Oxford University Press, Oxford, UK. pp. 232-263.

Thompson, J.D., Higgins, D.G., and Gibson, T.J. 1994. Clustal-W - Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22**(22): 4673-4680.

Turgeon, J., and Bernatchez, L. 2003. Reticulate evolution and phenotypic diversity in North American ciscoes, *Coregonus* ssp (Teleostei : Salmonidae): implications for the conservation of an evolutionary legacy. Conserv. Genet. **4**(1): 67-81.

Underhill, J.C. 1986. The fish fauna of the Laurentian Great Lakes Basin, the St. Lawrence lowlands, Newfoundland and Labrador. *In* The zoogeography of North American freshwater fishes *Edited by* C.H. Hocutt and E. O.Wiley. John Wiley & Sons, New York. pp. 105-136.

Wang, J. 2012. On the measurements of genetic differentiation among populations. Genet. Res. **94**(5): 275-289.

Witt, J.D.S., Zemlak, R.J., and Taylor, E.B. 2011. Phylogeography and the origins of range disjunctions in a north temperate fish, the pygmy whitefish (*Prosopium coulterii*), inferred from mitochondrial and nuclear DNA sequence analysis. J. Biogeogr. **38**(8): 1557-1569.

Wilson, C., and Hebert, P. 1998. Phylogeography and postglacial dispersal of lake trout (*Salvelinus namaycush*) in North America. Can. J. Fish. Aquat. Sci. **55**(4): 1010-1024.

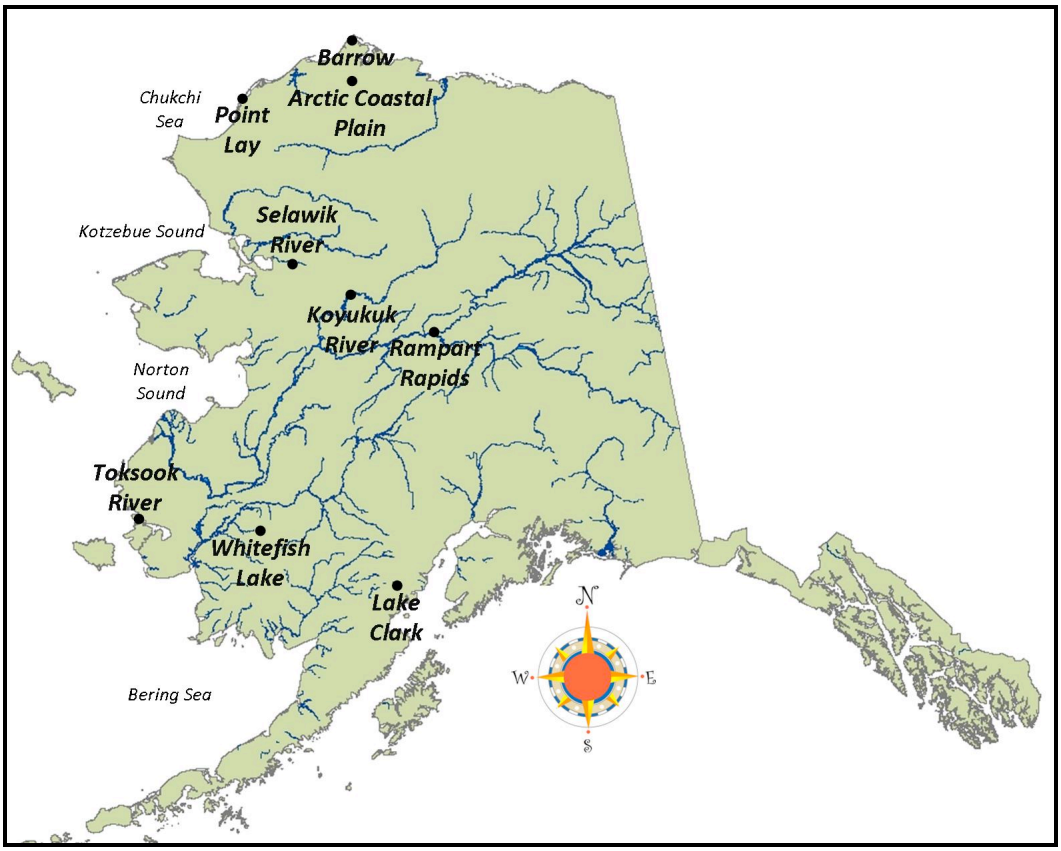

Fig. 2.1. Map of sampling locations.

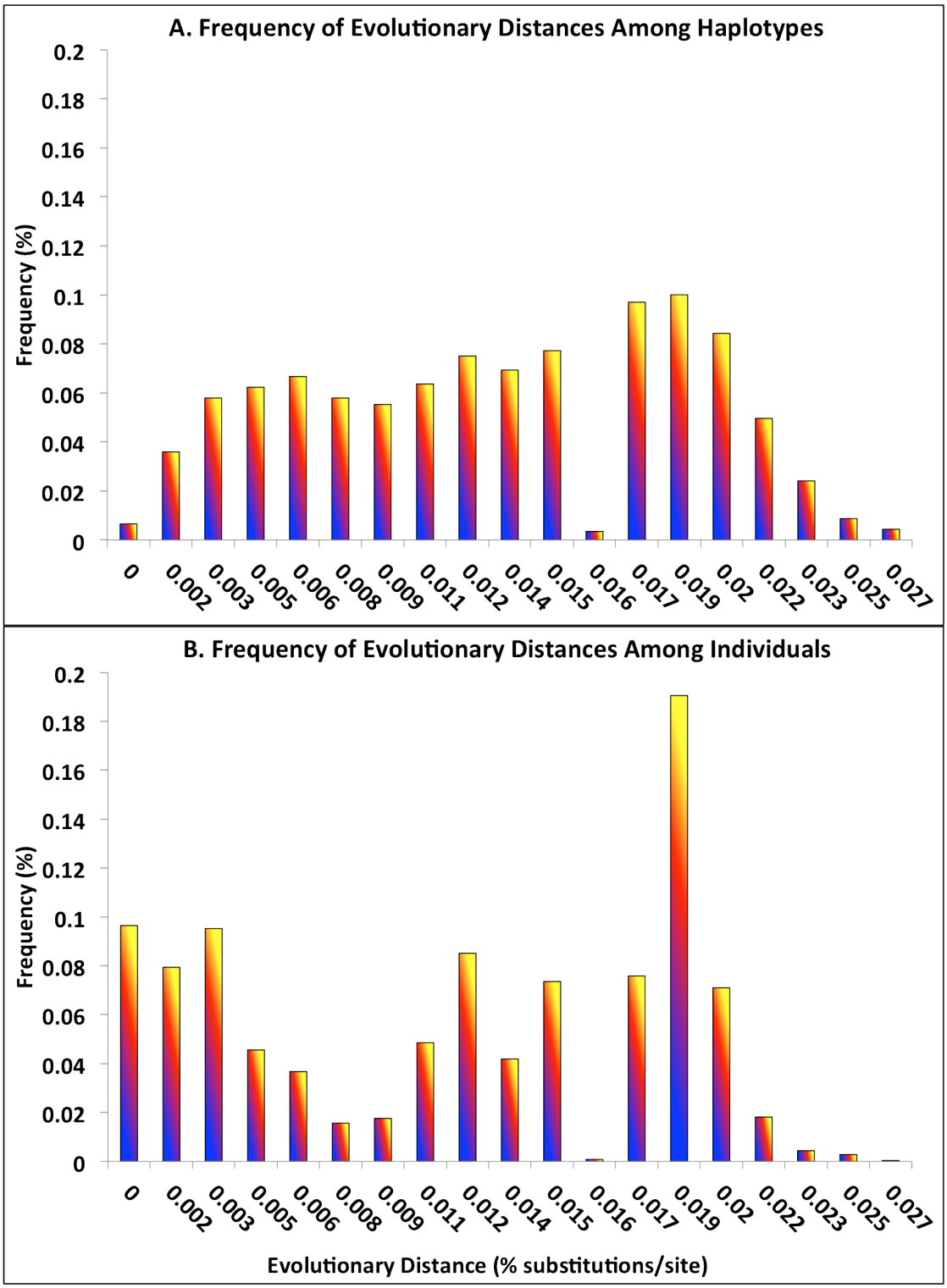

Fig. 2.2. Distribution of evolutionary distances: A) among haplotypes (n = 68); and B) among individuals (n = 305).

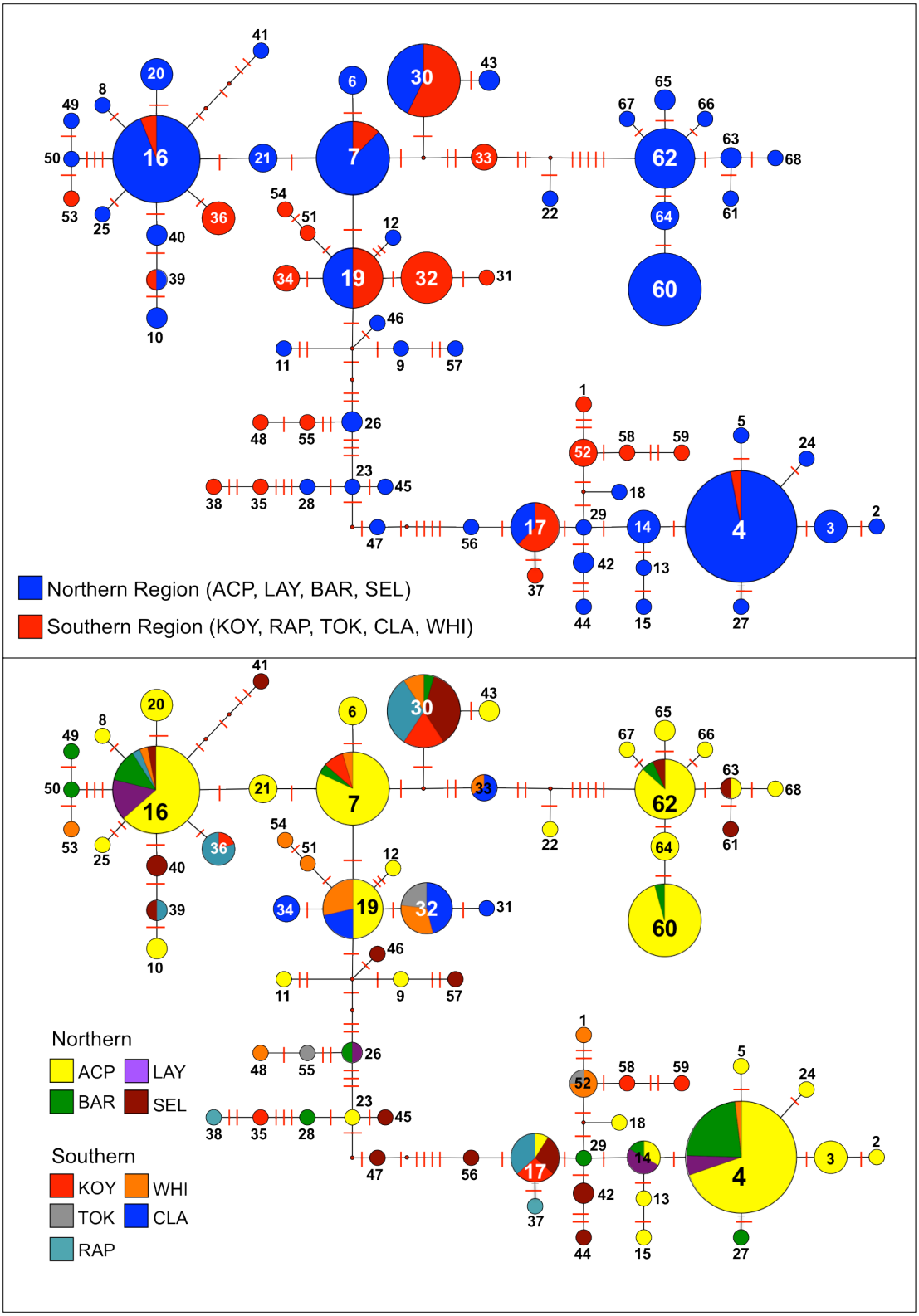

Fig. 2.3. Minimum spanning network of *C. sardinella* haplotypes.

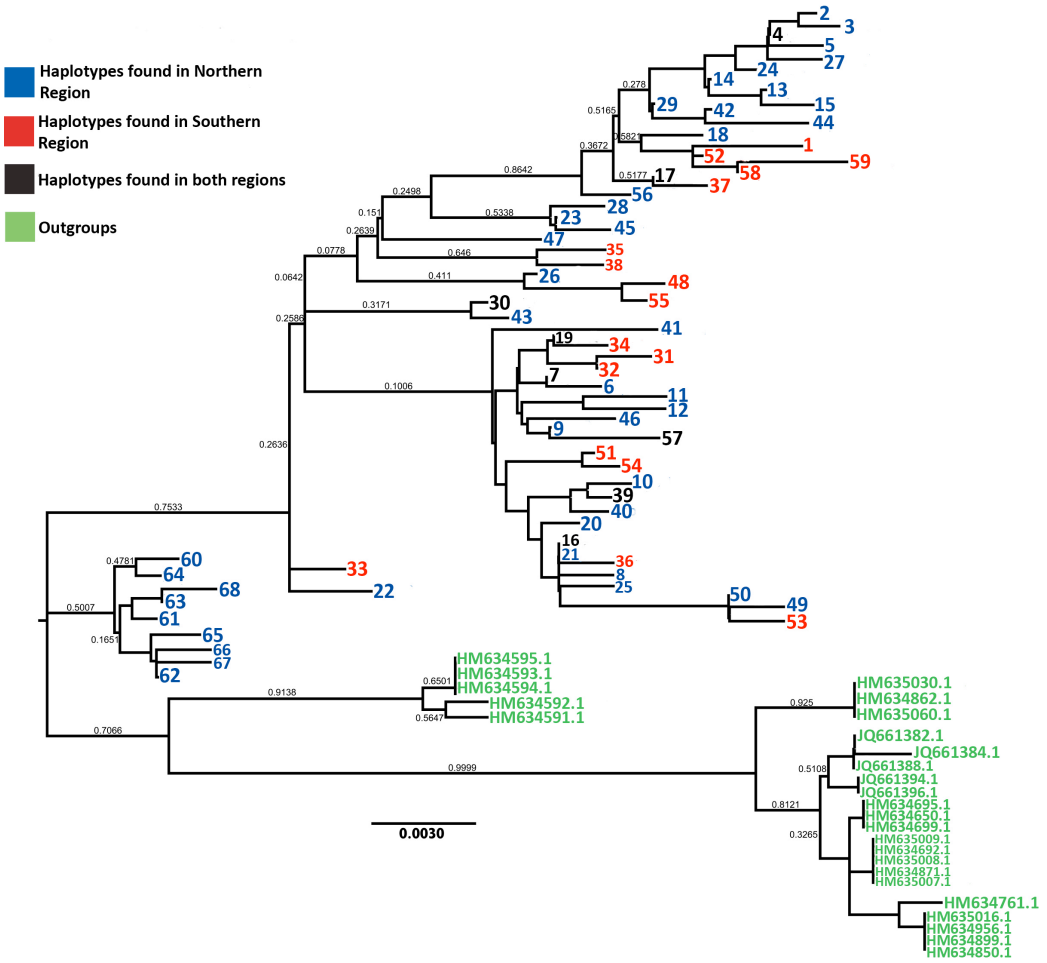

Fig. 2.4. Maximum likelihood phylogeny of the 68 unique haplotypes.

Table 2.1. Unique haplotypes identified in least cisco from northwestern Alaska.

	ACP	LAY	BAR	SEL	KOY	RAP	CLA	WHI	TOK
Hap_1	-	-	-	-	-	-	-	1	-
Hap_2	1	-	-	-	-	-	-	-	-
Hap_3	5	-	-	-	-	-	-	-	-
Hap_4	37	3	12	-	-	-	-	1	-
Hap_5	1	-	-	-	-	-	-	-	-
Hap_6	4	-	-	-	-	-	-	-	-
Hap_7	18	-	1	-	2	-	-	1	-
Hap_8	1	-	-	-	-	-	-	-	-
Hap_9	1	-	-	-	-	-	-	-	-
Hap_10	2	-	-	-	-	-	-	-	-
Hap_11	1	-	-	-	-	-	-	-	-
Hap_12	1	-	-	-	-	-	-	-	-
Hap_13	1	-	-	-	-	-	-	-	-
Hap_14	2	3	1	-	-	-	-	-	-
Hap_15	1	-	-	-	-	-	-	-	-
Hap_16	21	5	4	1	-	1		1	-
Hap_17	1	-	-	3	3	4	-	-	-
Hap_18	1	-	-	-	-	-	-	-	-
Hap_19	7	-	-	-	-	-	3	4	-
Hap_20	5	-	-	-	-	-	-	-	-
Hap_21	3	-	-	-	-	-	-	-	-
Hap_22	1	-	-	-	-	-	-	-	-
Hap_23	1	-	-	-	-	-	-	-	-
Hap_24	1	-	-	-	-	-	-	-	-
Hap_25	1	-	-	-	-	-	-	-	-
Hap_26	-	1	1	-	-	-	-	-	-
Hap_27	-	-	1	-	-	-	-	-	-
Hap_28	-	-	1	-	-	-	-	-	-
Hap_29	-	-	1	-	-	-	-	-	-
Hap_30	-	-	1	8	4	7	-	2	-
Hap_31	-	-	-	-	-	-	1	-	-
Hap_32	-	-	-	-	-	-	6	4	3
Hap_33	-	-	-	-	-	-	2	1	-
Hap_34	-	-	-	-	-	-	3	-	-
Hap_35	-	-	-	-	1	-	-	-	-
Hap_36	-	-	-	-	1	4	-	-	-
Hap_37	-	-	-	-	-	1	-	-	-
Hap_38	-	-	-	-	-	1	-	-	-
Hap_39	-	-	-	1	-	1	-	-	-
Hap_40	-	-	-	2	-	-	-	-	-
Hap_41	-	-	-	1	-	-	-	-	-
Hap_42	-	-	-	2	-	-	-	-	-
Hap_43	-	-	-	2	-	-	-	-	-
Hap_44	-	-	-	1	-	-	-	-	-
Hap_45	-	-	-	1	-	-	-	-	-
Hap_46	-	-	-	1	-	-	-	-	-
Hap_47	-	-	-	1	-	-	-	-	-

Table 2.1 continued

	ACP	LAY	BAR	SEL	KOY	RAP	CLA	WHI	TOK
Hap_48	-	-	-	-	-	-	-	1	-
Hap_49	-	-	1	-	-	-	-	-	-
Hap_50	-	-	1	-	-	-	-	-	-
Hap_51	-	-	-	-	-	-	-	1	-
Hap_52	-	-	-	-	-	-	-	3	1
Hap_53	-	-	-	-	-	-	-	1	-
Hap_54	-	-	-	-	-	-	-	1	-
Hap_55	-	-	-	-	-	-	-	-	1
Hap_56	-	-	-	1	-	-	-	-	-
Hap_57	-	-	-	1	-	-	-	-	-
Hap_58	-	-	-	-	1	-	-	-	-
Hap_59	-	-	-	-	1	-	-	-	-
Hap_60	22	-	1	-	-	-	-	-	-
Hap_61	-	-	-	1	-	-	-	-	-
Hap_62	13	-	1	1	-	-	-	-	-
Hap_63	1	-	-	1	-	-	-	-	-
Hap_64	4	-	-	-	-	-	-	-	-
Hap_65	2	-	-	-	-	-	-	-	-
Hap_66	1	-	-	-	-	-	-	-	-
Hap_67	1	-	-	-	-	-	-	-	-
Hap_68	1	-	-	-	-	-	-	-	-

Table 2.2. Sample sizes and locations for sites.

Site	n	Lat	Long
Arctic Coastal Plain	163	70.508	-155.367
Selawik River	29	66.604	-160.007
Point Lay	12	69.763	-163.053
Barrow	27	71.295	-156.256
Koyukuk River	13	64.88	-157.7
Rampart Rapids	19	65.535	-150.179
Lake Clark	15	60.214	-154.383
Whitefish Lake	22	61.37	-160.026
Toksook River	5	60.515	-164.971

Table 2.3. Descriptive statistics for sequences.

Site	Avg No. diff (*K*)	Hap Div (H_d)	No. Hap (h)	Max dis	Avg Gen Dis (SD)	Hap Richness
All	8.266	0.935	68	0.027	0.011 ± 0.007	40.769
ACP	8.379	0.894	32	0.023	0.012 ± 0.008	11.147
SEL	7.337	0.916	17	0.022	0.010 ± 0.006	12.566
LAY	7.606	0.758	4	0.019	0.010 ± 0.008	4.000
BAR	8.188	0.795	13	0.027	0.011 ± 0.009	9.651
KOY	7.487	0.872	7	0.023	0.011 ± 0.007	5.358
RAP	6.07	0.807	7	0.019	0.009 ± 0.007	4.718
CLA	2	0.79	5	0.009	0.003 ± 0.003	4.254
WHI	6.506	0.931	13	0.025	0.009 ± 0.006	6.926
TOK	6.4	0.7	3	0.019	0.009 ± 0.007	3.000

Table 2.4 Homogeneity test results among sites, corrected for multiple testing. $P <$ 0.0001 represented as *.

	ACP	LAY	BAR	SEL	KOY	RAP	CLA	WHI
LAY	**0.094 ± 0.010**							
BAR	**0.124 ± 0.025**	**0.524 ± 0.021**						
SEL	*	0.001 ± 0.001	*					
KOY	*	*	*	**0.422 ± 0.039**				
RAP	*	*	*	**0.278 ± 0.023**	**0.421 ± 0.015**			
CLA	*	*	*	*	*	*		
WHI	*	*	*	*	0.008 ± 0.003	*	**0.178 ± 0.006**	
TOK	*	0.002 ± 0.001	0.006 ± 0.003	0.014 ± 0.007	0.003 ± 0.001	*	**0.231 ± 0.017**	**0.794 ± 0.016**

Table 2.5. Φ_{ST} values indicating levels of genetic differentiation among sites above diagonal, corrected *p*-values below diagonal. $P < 0.0001$ represented as *.

	ACP	LAY	BAR	SEL	KOY	RAP	CLA	WHI	TOK
ACP		**0.059**	**0.024**	0.088	0.099	0.136	0.141	0.062	0.171
LAY	**0.011 ± 0.001**		**0.046**	0.142	0.185	0.198	0.225	0.123	0.265
BAR	**0.030 ± 0.001**	**0.096 ± 0.003**		0.130	0.156	0.182	0.207	0.110	0.238
SEL	*	*	*		**-0.004**	**0.012**	0.142	0.051	0.163
KOY	*	*	*	**0.459 ± 0.005**		**-0.021**	0.170	0.065	0.197
RAP	*	*	*	**0.202 ± 0.005**	**0.627 ± 0.005**		0.201	0.098	0.232
CLA	*	*	*	*	*	*		**0.025**	**0.011**
WHI	0.001 ± 0.0002	0.001 ± 0.0002	*	0.003 ± 0.001	0.010 ± 0.001	0.001 ± 0.0002	**0.140 ± 0.004**		**0.032**
TOK	0.003 ± 0.001	0.004 ± 0.001	0.001 ± 0.0002	0.001 ± 0.0002	0.005 ± 0.001	0.002 ± 0.001	**0.418 ± 0.005**	**0.251 ± 0.004**	

Table 2.6. A breakdown of characteristics of haplotypes 60-68.

Haplotype ID	Number of individuals with haplotype	Populations represented	Average genetic distance from haplotypes 1-59
Hap 60	23	ACP, BAR	0.016
Hap 61	1	SEL	0.016
Hap 62	15	ACP, BAR, SEL	0.016
Hap 63	2	ACP, SEL	0.016
Hap 64	4	ACP	0.016
Hap 65	2	ACP	0.017
Hap 66	1	ACP	0.017
Hap 67	1	ACP	0.018
Hap 68	1	ACP	0.018

Chapter 3: Population genetic structure of least cisco (*Coregonus sardinella*) from freshwater lakes on the Arctic Coastal Plain of Alaska[2]

Abstract

The least cisco (*Coregonus sardinella*), which is broadly distributed in freshwater lakes across the Arctic Coastal Plain of Alaska, encounter highly variable environmental conditions across a broad range of spatial and temporal scales. Little is known about these lake populations, including life history patterns and population genetic structure. We collected least cisco from seven Arctic Coastal Plain lakes and genotyped six microsatellite loci for each individual. Patterns of departure from Hardy-Weinberg equilibrium (HWE) were not consistent on a lake-by-lake or locus-by-locus basis. Of the six loci, only one locus (CLB129) provided reliable data that were then compared to mtDNA sequences from the same individuals. The Φ_{ST} value for mtDNA was higher than the analogous θ value for CLB129. The individual lakes of the Arctic Coastal Plain may not represent individual subpopulations as initially assumed. We suggest that historical migration and extinction/colonization dynamics were important factors shaping the genetic relationships among groups of least cisco. Ultimately, Alaskan least cisco may have functioned as a metapopulation historically, but present populations are too isolated to be considered a metapopulation today.

Introduction

Least cisco (*Coregonus sardinella*)

Least cisco exhibit high levels of phenotypic and life history plasticity and have adapted to extreme Arctic environmental conditions (Dillinger 1989). For example, they can be amphidromous, lacustrine, or riverine (Harper et al. 2007). Both normal and dwarf

[2] Padula, V. M., D. Causey, and J. A. Lopez. 2013. Chapter 3: Population genetic structure of least cisco (*Coregonus sardinella*) from freshwater lakes on the Arctic Coastal Plain of Alaska. Prepared for submission to the University of Alaska Fairbanks Graduate School. Formatted according to the Canadian Journal of Fisheries and Aquatic Sciences guidelines.

freshwater resident least cisco occur in Trout Lake and Peters Lake in the Yukon Territory of Canada (Dillinger 1989; Mann 1974). Anadromous riverine least cisco from the Mackenzie River are comparatively larger, heavier, longer-lived, and more fecund than fish found in Trout Lake and Peters Lake (Dillinger 1989; Mann 1974; Mann and McCart 1981). Although little to no data exist for least cisco inhabiting the Arctic Coastal Plain of Alaska, the diversity of habitats within this ecosystem can potentially support life history trait variation and genetic differentiation among least cisco groups in this region.

The Arctic Coastal Plain landscape

The northernmost region of the North Slope, the Arctic Coastal Plain, is 50,000 km^2, is bordered on the west and north by the Chukchi and Beaufort seas, and extends east nearly to the U.S.-Canada border. During the Quaternary period, a shallow sea deposited silt, sand, and gravel on the Arctic Coastal Plain ecosystem (Black 1964), which is covered by a complex hydrographic network subject to seasonal extremes such as hot, dry summers and extremely cold winters. This region is characterized by thick permafrost, with permafrost-related surface features such as pingos, ice wedge polygons, peat ridges, and frost boils. Hydrologically, the region experiences low annual precipitation, and the permafrost and low topographic gradient result in a landscape dominated by wetlands, ponds, and lakes.

Lakes are a major landscape feature and vary in depth and levels of connectivity to other hydrological features such as wetlands, streams, and rivers. Variation in lake depth results in variation in the year-round availability of liquid water within lakes. The smaller and shallower (< 2 m) lakes freeze to the bottom and are liquid only during the warm season, and upwelling is unlikely as permafrost exists below the surface of these lakes. Surface runoff for recharge is important for these lakes because they are subject to evaporative loss during summer (Miller et al. 1980). Although landlocked lakes are present, many lake systems are seasonally linked by shallow streams, which form complex hydrographic networks (Martin et al. 2009). Many major streams originate from the south, and lakes of thermokarst and non-thermokarst origin constitute over 14% of the landscape (Hobbie 1973; Jorgenson and Heiner 2003).

Lake ecosystems, especially ecosystems that have persisted a long of time, are natural systems in which to study the underlying mechanisms of speciation, colonization, adaptation and diversification (Cristescu et al. 2010). Physical and biotic processes operating on different spatial and temporal scales create a mosaic of aquatic habitats for fishes (Collins and Glenn 1997; Wu and Loucks 1995). During the spring break-up, flooding patterns can vary substantially across the landscape; ephemeral and temporary connections between major bodies of water can have profound consequences for fish distributions. Small, young-of-the-year fish may be able to migrate through these connections, potentially populating lakes that are otherwise isolated during the rest of the summer. This complex hydrographic network means various channels and barriers to gene flow may exist for Arctic fishes, but we do not know the degrees of connectivity and isolation. Additionally, climate change is apparent in the Arctic. Modern warming trends may alter the hydrographic network of the Alaskan Arctic, consequently altering fish distributions (Hershey et al. 2006) and gene flow among fish populations.

Study goals

Least cisco populations have been studied in multiple major river drainages along the Arctic Alaskan and Canadian coasts (Brown 2004; Brown 2009; Dillinger 1989; Hallberg 1989), but little or no prior research has been reported for least cisco populations inhabiting the inland freshwater lakes of the Arctic Coastal Plain. How are least cisco genetically related to each other on a fine landscape scale, such as among a group of closely located freshwater lakes on the Arctic Coastal Plain? Developing a better understanding of the population structuring of a highly mobile and phenotypically diverse species such as the least cisco will provide us with a foundation upon which we can begin to understand how other fishes use the aquatic habitats of the Arctic Coastal Plain. Consequently, the goals of this research chapter were to: (i) genotype least cisco from seven lakes on the Arctic Coastal Plain for six microsatellite loci: BWF2, Aut139, CLB107, CLB129, CLC4, CLC101; and (ii) compare indices of genetic diversity among lake populations.

Materials and Methods

Study site and sampling

The freshwater lakes from which least cisco were collected are located on the National Petroleum Reserve of Alaska, north of the Brooks Mountain Range (Fig. 3.1), and the unique identifiers of the lakes are related to their locations in survey plots mapped out during concurrent scientific investigations. Fish were collected from seven lakes in 2009 and 2010 were used in this study: lake 128 (n = 30), 161 (n = 20), 216 (n = 21), 276 (n = 16), 306 (n = 19), 566 (n = 10), and 567 (n = 14). Fish were caught with gill net, fyke net, or beach seines; and a pelvic or caudal fin from each fish was removed and preserved. Lake characteristics (number of samples collected, latitude, longitude, surface area, volume, and % unfrozen in April) are presented in Table 3.1.

Microsatellite analysis

Total DNA was extracted from tissue samples with a Puregene DNA Extraction Core Kit (Qiagen, Germantown, MD), following the manufacturer's protocol. A total of 76 microsatellite markers were tested on a small subset of least cisco DNA for use in this study. After these trials, all 2009 samples and a subsample from 2010 representing all lakes were analyzed at six microsatellite loci: BWF2 (Patton et al. 1997), Aut139 (Ramey et al. 2008), CLC7, CLB129, CLC4, and CLC101 (Schlei, FWS) (Table 3.2). Loci were amplified individually by the polymerase chain reaction (PCR). The markers were individually labeled with HEX™, PET™, VIC™, or 6-FAM™ dye (Table 3.2). Conditions for a 10 μL reaction volume were as follows: 0.5 μL DNA template (concentration of approximately 100 ng/mL), 1X GoTaq buffer (Promega Corporation, Madison, WI), 2 mM dNTPs, 2.5 mM $MgCl_2$, 1 μM forward and reverse primer, and 0.04 U/μL GoTaq (Promega Corporation, Madison, WI). Reactions were run for 40 cycles with an annealing temperature of 50°C. PCR products were checked by gel electrophoresis to verify amplifications of the correct product on a 2.5% agarose gel and 1 μL of each confirmed PCR product was diluted with 9 μL of Hi-Di Formamide™ (Applied Biosystems, Grand Island, NY). PCR products were sent to the Yale DNA Analysis Facility for fragment size analysis on an automated DNA sequencer (ABI 3100),

using ROX-500 or LIZ-500 size standard. Chromatograms were loaded into GeneMarker (http://www.softgenetics.com/GeneMarker.html) to score genotypes. Two researchers independently scored all individuals to reduce bias caused by scoring errors and difficulties in amplifying DNA, and a third researcher scored a subsample of individuals.

Statistical analysis

Basic descriptive statistics were calculated for each locus on a lake-by-lake basis. This included the number of unique genotypes, observed heterozygosity (H_O) and expected heterozygosity (H_E), effective number of alleles (n_{eff}), and F_{IS}. H_O and H_E were calculated at each locus in GENEPOP version 4.2 (Raymond and Rousset 1995; Rousset 2008) and used to test whether loci were in Hardy-Weinberg equilibrium (HWE). Sequential Bonferroni methods were implemented to correct against type I errors (Rice 1989). Values of n_{eff} and F_{IS} were estimated using Microsoft Excel (2011), using the equations $n_{eff} = 1/(1-H_E)$ and $F_{IS} = 1- (H_O/H_E)$.

Pseudoexact *G* tests for homogeneity were then performed in ARLEQUIN version 3.5.1.3 (Excoffier and Lischer 2010), with the data grouped into five hierarchical structures. These tests were used to determine if differences in genetic compositions existed across groups. First the data were tested for each of the seven lakes, assuming they each represented unique populations. The data were then grouped into three geographic regions: (i) lakes 128 and 161; (ii) lakes 216, 276, and 306; and (iii) lakes 566 and 567. Homogeneity tests were then performed within each region.

At this point we determined that CLB129 was the only locus that provided reliable and informative data for this study (see results). Therefore, using only CLB129 data grouped into three regions, we performed an analysis of molecular variance (AMOVA) (Excoffier et al. 1992) in ARLEQUIN version 3.5.1.3 (Excoffier and Lischer 2010) and compared those results to those of mtDNA sequences used in the previous research chapter. From the original sample set of sequences, we took sequences from individuals specifically collected from the seven lakes used in this study, and grouped those sequences into the same three regions used here (Table 3.5). A pseudoexact *G* test and AMOVA were performed on the mtDNA sequences.

Results

Microsatellite scoring

Although a standard method for genotyping quality control has not been definitively established, some measures have been suggested for checking the quality of genotypic data. These include: (i) once loci have been scored, rescore a subset of genotypes to calculate error rate and ensure that scoring is accurate; (ii) ensure that all alleles are being amplified, test for homozygote excess patterns that are consistent with null alleles or large allele dropout; (iii) test data for linkage equilibrium between alleles at different loci; (iv) to ensure selective neutrality, test for outlier loci; and (v) perform defined crosses when possible to ensure Mendelian inheritance (Selkoe and Toonen 2006).

We encountered some difficulties in scoring loci, which were only apparent after we had analyzed a large number of individuals. For example, the BWF2 primer may have been amplifying two sections of the least cisco genome because one or two peaks were appearing within the previously published allelic range (Patton et al. 1997), but additional peaks were appearing outside of the previously published range. In roughly half the chromatograms, the two sets of peaks were distant from each other and could be scored. In the other chromatograms, however, but the two sets of peaks were so close in the remaining reports that they could not be scored reliably. We also encountered difficulties in amplifying DNA for CLC7. Figure 3.2 depicts the reductions in usable data, the high blue bars versus the low orange bars highlight the particular difficulties we encountered with BWF2 and CLC7.

Descriptive statistics

Descriptive statistics are reported in Table 3.3. Information for CLC7 and BWF2 were included, although those loci had few data. Bold face type for H_O and H_E values indicated a significant p-value for $\boldsymbol{X}^2$ tests, indicating a departure from HWE. Highly aberrant F_{IS} values ($F_{IS} < -0.1$ or $F_{IS} > 0.1$) were also in bold and were indicative of departure from HWE. On a lake-by-lake basis, the only locus that conformed to HWE

was CLB129. The other loci either lacked sufficient data or departed from HWE at certain lakes.

CLB129 vs. mtDNA

When the three-region data for CLB129 were compared to the three-region data for mtDNA, both DNA types had significant *P*-values for homogeneity tests, indicating that genetic compositions varied across the three regions at the nuclear and mitochondrial DNA level (Table 3.6). However, F_{ST} values calculated from the AMOVA differ between the DNA types. F_{ST} was an order of magnitude greater and significant for mtDNA, while F_{ST} was not significant for CLB129 (Table 3.6).

Discussion

Multilocus allele frequency data provide information about population of origin for individuals and how many populations are found in a specific region (Selkoe and Toonen 2006). Highly polymorphic sequence or microsatellite data provide information about a population's recent evolutionary history, such as whether it expanded or contracted in the past, and how the sizes of past populations compare to the sizes of current populations (Selkoe and Toonen 2006). Multilocus genotype data provide information about the genetic relationships among individuals and whether individuals have moved (Selkoe and Toonen 2006).

We encountered a number of difficulties with the analyses of our microsatellite data, and these difficulties likely stemmed from several conditions we did not foresee at the outset of this study. On a lake-by-lake and locus-by-locus basis, χ^2 tests and F_{IS} values indicated inconsistent patterns of departure from HWE. In some cases, certain loci did not successfully amplify for individuals from specific lakes (e.g. CLC101, CLC7, and CLC4 for lake 567), which further limited data availability for statistical analyses. In light of these data limitations, we chose to only use data from one microsatellite locus, CLB129, and compare these data to mtDNA sequence data collected in the previous research chapter. We ran these analyses with the data grouped into three regions. Group one comprised lakes 128 and 161, group two comprised lakes 216, 276, and 306, and group three comprised lakes 566 and 567. Homogeneity tests among groups were

significant for both types of DNA, which indicated that AMOVA analyses could be run. However, the AMOVAs yielded very different results for CLB129 and mtDNA. F_{ST} for CLB129 was 0.00624 and not significant, while Φ_{ST} for mtDNA was 0.08167 and significant.

Even though the same individuals were used in the CLB129 and mtDNA analyses, several factors may explain why F_{ST} values differ so greatly: (i) small sample size, especially considering only one microsatellite locus was used; (ii) different mutations rates between mtDNA and microsatellite DNA; and (iii) different modes of inheritance. Microsatellite DNA is diploid and inherited from both parents, while mtDNA is haploid and maternally inherited. Maternal inheritance reduces N_e by one-fourth. This value, accompanied with migrations rate (m), is part of the migration-drift equilibrium and is used in calculating F_{ST} ($F_{ST} = 1/(4N_em + 1)$. A small N_e is more affected by genetic drift, but migration can counteract the effects of drift because new alleles may be introduced to populations through migrants. However, insufficient data limit the confirmation of any of these speculations.

Future landscape genetics studies of least cisco

Least cisco are highly mobile and inhabit a hydrographic landscape in which water bodies are intricately connected, and we suggest that historical migration, extinction, and colonization dynamics were important factors shaping the genetic relationships among groups of least cisco. Ultimately, Alaskan least cisco may have functioned as a metapopulation historically, but present populations are too isolated to be considered a metapopulation. An investigation of least cisco population genetics accompanied with an investigation of their seasonal distribution in northern Alaska would create a more informed picture of how these highly mobile species are utilizing the landscape. For fish that are highly mobile, such as the least cisco, surface level connectivity may shape the population genetic structure by allowing corridors for migration or barriers to gene flow. Additionally, climate change might affect surface water connectivity in the future, with unknown consequences to fish populations. The broad distribution of least cisco makes them a good indicator species of what might be

occurring to other fish populations on the Arctic Coastal Plain.

Changes in overall water balance, timing and magnitude of seasonal water and energy fluxes resulting from climate change could significantly alter the aquatic and semi-aquatic habitats of the Arctic Coastal Plain (Martin et al. 2009). Drier conditions are already being detected on the Coastal Plain as precipitation fails to compensate for evapotranspiration (Hinzman et al. 2005), a process that plays a key role in Arctic hydrology, and becomes stronger as temperatures rise (Bowling et al. 2003; Kane 1997). As the rate of evapotranspiration can be limited by water or energy availability, the degree to which evapotranspiration affects hydrology will vary seasonally. The effects are expected to be most strongly felt in summer, when drying will occur earlier in the season and to a greater degree. Energy is limited in the fall and water is limited in the winter, so evapotranspiration will not have as great an effect in those seasons (Martin et al. 2009). Winter conditions will be more affected by changes in snow cover, consequently impacting Arctic lake heat budgets because increased snow cover will provide more insulation and thereby thin ice cover (Schindler and Smol 2006). As winter turns to spring, the thinner ice melts more quickly, especially in warmer air temperatures, causing earlier ice-out and rises in water temperature (Martin et al. 2009). These changes in water flow, water temperature and air temperature will ultimately affect individual fitness and population resilience of Arctic organisms (Svenning and Gullestad 2002).

The process through which the data were compiled in this study provided valuable insights for future research. A number of reasons may underlie the HWE departures and inconsistent F_{ST} patterns seen in our data. Insufficient data, coupled with poor data quality are the most likely reasons. Microsatellite markers, and other nuclear DNA loci, should be specifically designed for least cisco so that primers could be optimized much more easily during the PCR process. The original design of this study, which assumed a rather static view of the subpopulation structuring of least cisco, should also be altered. Individual lakes may not represent individual subpopulations as initially assumed, and more dynamic dispersal processes may be at work. Samples should also be collected from more of the smaller lakes on the Arctic Coastal Plain as well as from larger water bodies

such as Teshekpuk Lake, Smith Bay, Admiralty Bay, Elson Lagoon, Iko Bay, Peard Bay, and Wainwright Inlet. Additionally, resampling of sites in different seasons and spawning aggregations will provide information on population admixture during migration periods. Other variables such as the age and sex of each fish caught and used in the genetic analysis might provide insight about generation overlaps and sex ratios.

Acknowledgments

Samples from ACP were collected in 2009 and 2010 from the freshwater lakes on the National Petroleum Reserve of Alaska (NPR-A) under the permit: [148907-1] The distribution and structure of fish communities in arctic lakes on the North Slope, Alaska.). My gratitude goes to the numerous field technicians and research assistants who collected fin clips during those field seasons. My gratitude also goes to Patrick Barry for reviewing my data and assisting me with analysis.

References

Black, R.F. 1964. Gubik formation of Quaternary age in northern Alaska: Paper 302-C. U.S. Geological Survey. p. 59-91.

Bowling, L.C., Kane, D.L., Gieck, R.E., Hinzman, L.D., and Lettenmaier, D.P. 2003. The role of surface storage in a low-gradient Arctic watershed. Water Resour. Res. **39**(4).

Brown, R.J. 2004. A biological assessment of whitefish species harvested during the spring and fall in the Selawik River Delta, Selawik National Wildlife Refuge, Alaska: Alaska Fisheries Technical Report Number 77. U.S. Fish and Wildlife Service, Fairbanks Fish and Wildlife Field Office. 50pp.

Brown, R.J. 2009. Distributions and demographics of whitefish species in the Upper Koyukuk River Drainage, Alaska, with emphasis on seasonal migrations and important habitats of broad whitefish and humpback whitefish: Alaska Fisheries Technical Report Number 104. U.S. Fish and Wildlife Service, Fairbanks Fish and Wildlife Field Office. 24pp.

Collins, S.L., and Glenn, S.M. 1997. Effects of organismal and distance scaling on analysis of species distribution and abundance Ecol. Appl. **7**(2): 543-551.

Cristescu, M.E., Adamowicz, S.J., Vaillant, J.J., and Haffner, D.G. 2010. Ancient lakes revisited: from the ecology to the genetics of speciation. Mol. Ecol. **19**(22): 4837-4851.

Dillinger, R.E. 1989. An Analysis of Taxonomic Status of the *Coregonus autumnalis* Species Complex in North America, and an Investigation of the Life History of Whitefishes and Ciscoes (Pisces; Coregoninae) in North America and Eurasia, Memorial University of Newfoundland, St. John's, Newfoundland, Canada. 163pp.

Excoffier, L., and Lischer, H.E.L. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol. Ecol. Resour. **10**(3): 564-567.

Excoffier, L., Smouse, P.E., and Quattro, J.M. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes - Application to human mitochondrial-DNA restriction data. Genetics **131**(2): 479-491.

Hallberg, J.E. 1989. Abundance and size composition of Chatanika River least cisco and humpback whitefish with estimates of exploitation by recreational fishermen:

Fishery Data Series Report Number 108. Alaska Department of Fish and Game. 22pp.

Harper, K.C., Harris, F., Brown, R.J., Wyatt, T., and Cannon, D. 2007. Stock assessment of broad whitefish, humpback whitefish and least cisco in Whitefish Lake, Yukon Delta National Wildlife Refuge, Alaska, 2001-2003. Alaska Fisheries Technical Report Number 88. U.S. Fish and Wildlife Service, Fairbanks Fish and Wildlife Field Office. 55pp.

Hershey, A.E., Beaty, S., Fortino, K., Keyse, M., Mou, P.P., O'Brien, W.J., Ulseth, A.J., Gettel, G.A., Lieniesch, P.W., Lueke, C., McDonald, M.E., Mayer, C.H., Miller, M.C., Richards, C., Schuldt, J.A., and Whalen, S.C. 2006. Effect of landscape factors on fish distribution in Arctic Alaskan lakes. Freshwater Biol. **51**(1): 39-55.

Hinzman, L.D., Bettez, N.D., Bolton, W.R., Chapin, F.S., Dyurgerov, M.B., Fastie, C.L., Griffith, B., Hollister, R.D., Hope, A., Huntington, H.P., Jensen, A.M., Jia, G.J., Jorgenson, T., Kane, D.L., Klein, D.R., Kofinas, G., Lynch, A.H., Lloyd, A.H., McGuire, A.D., Nelson, F.E., Oechel, W.C., Osterkamp, T.E., Racine, C.H., Romanovsky, V.E., Stone, R.S., Stow, D.A., Sturm, M., Tweedie, C.E., Vourlitis, G.L., Walker, M.D., Walker, D.A., Webber, P.J., Welker, J.M., Winker, K., and Yoshikawa, K. 2005. Evidence and implications of recent climate change in northern Alaska and other Arctic regions. Climatic Change **72**(3): 251-298.

Hobbie, J.E. 1973. Arctic Limnology: A Review. *In* Alaskan Arctic tundra. Proceedings of the 25th Anniversary Celebration of the Naval Arctic Research Laboratory. *Edited by* M.E. Britton. pp. 127-168.

Jorgenson, M.T., and Heiner, M. 2003. Ecosystems of Northern Alaska, Unpublished 1:2.5 million-scale map. ABR, Inc., Fairbanks, AK and The Nature Conservancy, Anchorage, AK.

Kane, D.L. 1997. The impact of hydrologic perturbations on Arctic ecosystems induced by climate change. *In* Global Change and Arctic Terrestrial Ecosystems. *Edited by* W.C. Oechel, T. Callaghan, T. Gilmanov, J.I. Holten, B. Maxwell, U. Molau and B. Sveinbjornsson. Springer Verlag Inc., New York. pp. 63-81.

Mann, G.J. 1974. Life history types of the least cisco (*Coregonus sardinella* Valenciennes) in Beaufort Sea drainages in the Yukon Territory, North Slope, and east Mackenzie River drainages. M.S. Thesis, University of Alberta, Edmonton, Alberta, Canada. 157pp.

Mann, G.J., and McCart, P.J. 1981. Comparison of sympatric dwarf and normal-populations of least cisco (*Coregonus sardinella*) inhabiting Trout Lake, Yukon Territory. Can. J. Fish. Aquat. Sci. **38**(2): 240-244.

Martin, P.D., Jenkins, J.L., Adams, F.J., Jorgenson, M.T., Matz, A.C., Payer, D.C., Reynolds, P.E., Tidwell, A.C., and Zelenak, J.R. 2009. Wildlife Response to Environmental Arctic Change: Predicting Future Habitats of Arctic Alaska. Report of the Wildlife Response to Environmental Arctic Change (WildREACH): Predicting Future Habitats of Arctic Alaska Workshop, 17-18 November 2008. Fish and Wildlife Service, Fairbanks, Alaska: U.S. 138pp.

Miller, M.C., Prentki, R.T., and Barsdate, R.J. 1980. Physics. *In* Limnology of Tundra Ponds. *Edited by* J.E. Hobbie. Dowden, Hutchinson, and Ross, Inc., Stroudsburg, PA. pp. 51-75.

Patton, J.C., Gallaway, B.J., Fechhelm, R.G., and Cronin, M.A. 1997. Genetic variation of microsatellite and mitochondrial DNA markers in broad whitefish (*Coregonus nasus*) in the Colville and Sagavanirktok rivers in northern Alaska. Can. J. Fish. Aquat. Sci. **54**(7): 1548-1556.

Ramey, A., Graziano, S.L., and Nielsen, J.L. 2008. Isolation and characterization of microsatellite loci from the Arctic cisco (*Coregonus autumnalis*). Mol. Ecol. Resour. **8**(2): 357-359.

Raymond, M., and Rousset, F. 1995. Genepop (Version-1.2) - Population-genetics software for exact tests and ecumenicism. J. Hered. **86**(3): 248-249.

Rice, W.R. 1989. Analyzing tables of statistical tests. Evolution **43**(1): 223-225.

Rousset, F. 2008. GENEPOP ' 007: a complete re-implementation of the GENEPOP software for Windows and Linux. Mol. Ecol. Resour. **8**(1): 103-106.

Schindler, D.W., and Smol, J.P. 2006. Cumulative effects of climate warming and other human activities on freshwaters of Arctic and subarctic North America. Ambio **35**(4): 160-168.

Selkoe, K.A., and Toonen, R.J. 2006. Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. Ecol. Lett. **9**(5): 615-629.

Svenning, M.A., and Gullestad, N. 2002. Adaptations to stochastic environmental variations: the effects of seasonal temperatures on the migratory window of Svalbard Arctic charr. Environ. Biol. Fish. **64**(1-3): 165-174.

Wu, J.G., and Loucks, O.L. 1995. From balance of nature to hierarchical patch dynamics: A paradigm shift in ecology. Q. Rev. Biol. **70**(4): 439-466.

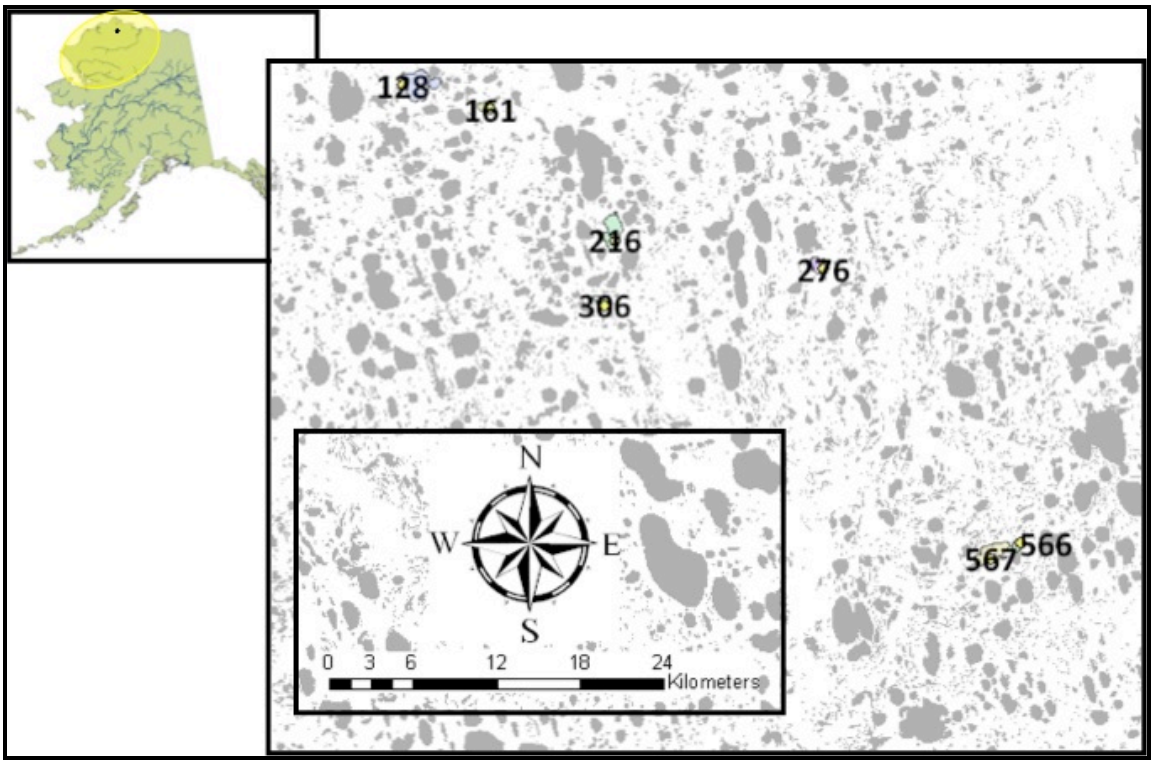

Fig. 3.1. Lakes from which least cisco were collected.

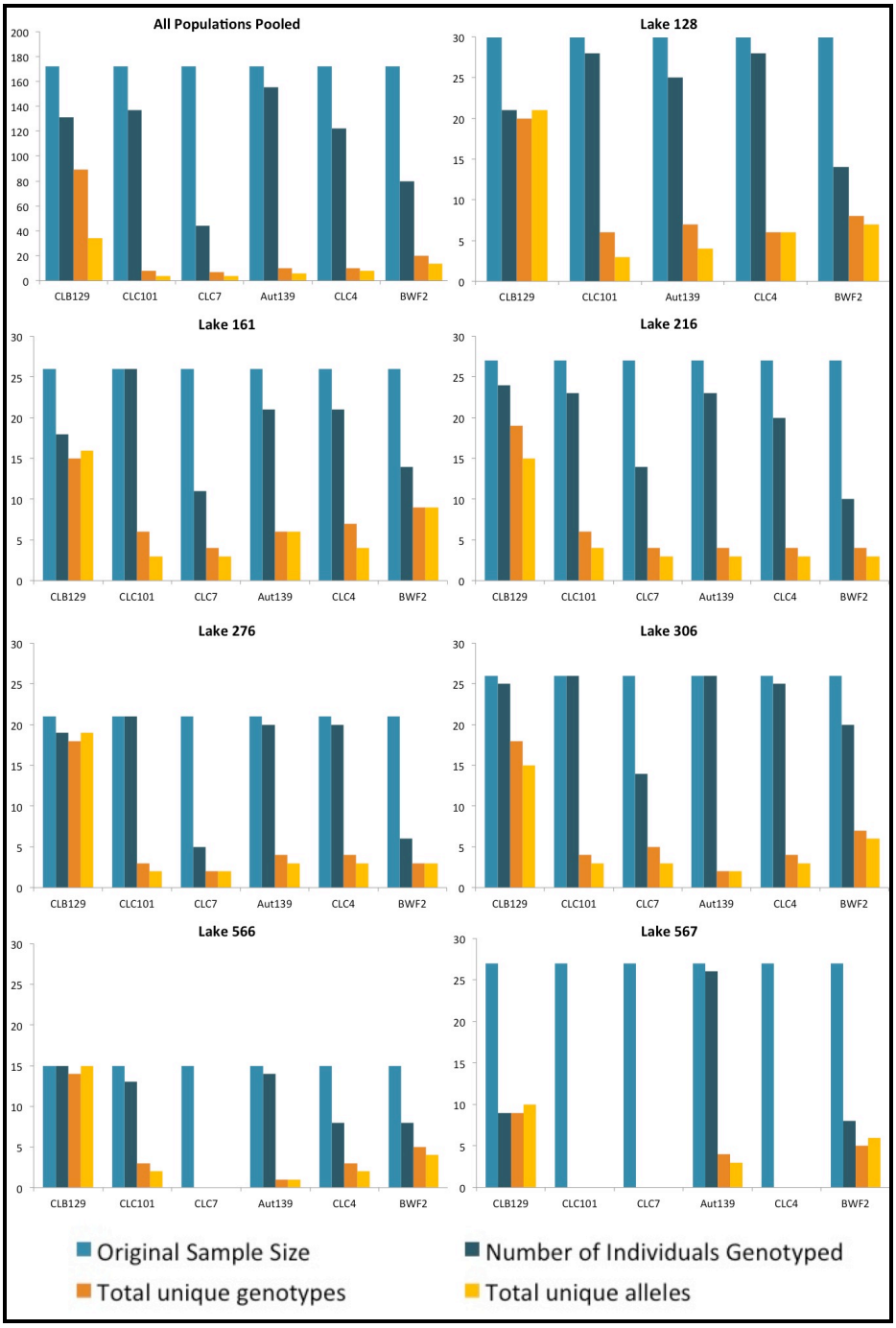

Fig. 3.2. Microsatellite sample size histograms.

Table 3.1. Number of samples collected, location, surface area, volume and % unfrozen.

Lake	n	Latitude (N)	Longitude (W)	Surface Area (m^2)	Volume (m^3)	% Unfrozen in April
128	30	70.7761	155.64038	3,636,109.73	6,026,782.78	could not be est. from SAR
161	20	70.772529	155.502532	593,398.61	669,420.01	could not be est. from SAR
216	21	70.694496	155.250717	2,465,118.73	4,753,278.87	75.6
276	16	70.672782	154.858108	1,128,528.77	806,132.60	57.1
306	19	70.646713	155.273345	571,342.43	1,258,718.48	89.9
566	10	70.488571	154.477798	435,748.47	753,248.35	87.7
567	14	70.484199	154.524398	2,357,276.02	2,635,138.93	64.2

Table 3.2. Description of microsatellite markers used in this study.

Name	Primer Sequence (5'-3')	Species	T_a (°C)	Reference
Aut139	F: GGG TAA AGC AGA ATG AC R: GGG AGT GTG TTT GTC T	*C. autumnalis*	50	Ramey *et al* 2008
BWF2	F: CGG ATA CAT CGG CAA CCT CTG R: AGA CAG TCC CCA ATG AGA AAA	*C. nasus*	50	Patton *et al* 1997
CLC4	F: CTC GAT CAT TGA TTG GTA AC*PET R: AAT GCA AGT ATG GGT ATT CAG A	*C. laurettae*	50	Schlei, FWS
CLC7	F: CCT CGT TGA TCT TAT AGA TGT C*VIC R: CCA ACC TAA GTG TAT GTA AAC C	*C. laurettae*	50	Schlei, FWS
CLB129	F: GCT TGG GAG ATT GGA GTC*PET R: GCC TGG TCT TTA TGA ATG C	*C. laurettae*	50	Schlei, FWS
CLC101	F: TTA TCA GGA CGT AAC CAT CAT C*6-FAM R: GGA TTA AAT GTC AGG AAT TGT G	*C. laurettae*	50	Schlei, FWS

Table 3.3. Descriptive statistics for microsatellite loci by site.

Lake ID		CLB129	CLC101	CLC7	Aut139	CLC4	BWF2
128	*n*	21	28	0	25	28	14
	n_a	21	3	0	4	6	7
n unique genotypes		20	6	0	7	6	8
	H_O	0.905	0.429	0.000	**0.360**	0.571	**0.214**
	H_E	0.956	0.532	0.000	**0.561**	0.572	**0.778**
	n_{eff}	22.660	2.139	1.000	2.277	2.337	4.500
	F_{is}	0.053	**0.195**	N/A	**0.358**	0.001	**0.724**
161	*n*	18	26	11	21	21	14
	n_a	16	3	3	6	4	9
n unique genotypes		15	6	4	6	7	9
	H_O	0.889	**0.269**	0.636	0.333	0.476	**0.429**
	H_E	0.924	**0.588**	0.558	0.345	0.632	**0.825**
	n_{eff}	13.125	2.429	2.265	1.527	2.716	5.727
	F_{is}	0.038	**0.542**	-0.140	0.034	**0.246**	**0.481**
216	*n*	24	23	14	23	20	10
	n_a	15	4	3	3	3	3
n unique genotypes		19	6	4	4	4	4
	H_O	0.875	0.609	0.643	0.435	0.500	0.500
	H_E	0.909	0.529	0.521	0.428	0.529	0.511
	n_{eff}	10.952	2.125	2.088	1.748	2.125	2.043
	F_{is}	0.037	**-0.150**	**-0.234**	-0.016	0.056	0.021
276	*n*	19	21	5	20	20	6
	n_a	19	2	2	3	3	3
n unique genotypes		18	3	2	4	4	3
	H_O	0.947	0.333	0.800	0.400	0.450	0.667
	H_E	0.950	0.396	0.533	0.412	0.499	0.545
	n_{eff}	20.084	1.656	2.143	1.699	1.995	2.200
	F_{is}	0.003	**0.158**	**-0.500**	0.028	0.098	**-0.222**
306	*n*	25	26	14	26	25	20
	n_a	15	3	3	2	3	6
n unique genotypes		18	4	5	2	4	7
	H_O	0.880	0.538	0.571	0.077	0.440	0.400
	H_E	0.874	0.456	0.635	0.075	0.474	0.487
	n_{eff}	7.955	1.839	2.739	1.082	1.902	1.950
	F_{is}	-0.007	**-0.180**	0.100	-0.020	0.072	**0.179**
566	*n*	15	13	0	14	8	8
	n_a	15	2	0	1	2	4
n unique genotypes		14	3	0	1	3	5
	H_O	0.933	0.462	0.000	0.000	0.250	**0.250**
	H_E	0.949	0.443	0.000	0.000	0.500	**0.675**
	n_{eff}	19.775	1.796	1.000	1.000	2.000	3.077
	F_{is}	0.017	-0.042	N/A	N/A	**0.500**	**0.630**
567	*n*	9	0	0	26	0	8
	n_a	10	0	0	3	0	6
n unique genotypes		9	0	0	4	0	5
	H_O	0.889	0.000	0.000	**0.231**	0.000	**0.250**
	H_E	0.876	0.000	0.000	**0.411**	0.000	**0.683**
	n_{eff}	8.053	1.000	1.000	1.698	1.000	3.158
	F_{is}	-0.015	N/A	N/A	**0.439**	N/A	**0.634**

Table 3.4. Tests of homogeneity for all loci.

	Number of populations	*P*-value
CLB129	**Genic Test**	
All lakes	7	***P*< 0.000**
3 regions	3	***P*< 0.000**
128/161	2	0.27644 ±0.00973
216/276/306	3	***P*< 0.000**
566/567	2	0.08261 ±0.00497
BWF2	**Genotypic Test**	
All lakes	7	NA
3 regions	3	***P*< 0.000**
128/161	2	0.17592 ±0.00529
216/276/306	3	***P*< 0.000**
566/567	2	0.81196 ±0.00344
CLC101		
All lakes	7	**0.00267 ±0.00155**
3 regions	3	**0.00206 ±0.00068**
128/161	2	0.85241 ±0.00258
216/276/306	3	**0.02457 ±0.00259**
566/567	2	NA
Aut139		
All lakes	7	NA
3 regions	3	**0.01359 ±0.00276**
128/161	2	0.06047 ±0.00297
216/276/306	3	**0.0153 ±0.00116**
566/567	2	**0.01372 ±0.00092**
CLC4		
All lakes	7	**0.00291 ±0.00113**
3 regions	3	0.07026 ±0.00600
128/161	2	0.18207 ±0.00567
216/276/306	3	**0.00231 ±0.00061**
566/567	2	NA
CLC7		
All lakes	7	0.2359 ±0.00949
3 regions	3	0.095 ± 0.00307
128/161	2	NA
216/276/306	3	0.32515 ± 0.00646
566/567	2	NA

Table 3.5. mtDNA sequences used for comparison to CLB129.

Region	n	n Haplotypes
128/161	29	15
216/276/306	23	7
566/567	14	7

Table 3.6. Comparison of CLB129 and mtDNA data for three regions.

3 Regions	*P* (homogeneity)	F_{ST}	*P* (F_{ST})
CLB129	***P* < 0.000**	0.00624	0.48944 ± 0.00159
mtDNA	***P* < 0.000**	0.08167	**0.00017 ± 0.00004**

General conclusions

In the first research chapter, a combination of analyses helped build a detailed interpretation of the evolutionary history of least cisco (*Coregonus sardinella*) in Alaska. This is the first report of a phylogeographic analysis of least cisco in Alaska, and they suggest that the least cisco's evolutionary history is different from other *Coregonus* species in Alaska. It showed that least cisco are relatively diverse across Alaska, with 68 unique haplotypes found in 305 individuals. F_{ST} values indicate incipient genetic differentiation, especially between the Arctic Ocean Drainage and the Bering Sea Drainage groups. The haplotype network and phylogeny show little evidence of geographic segregation among haplotypes, suggesting high levels of gene flow. But some haplotype groups in the network are restricted to a region, although it is not possible to draw a one to one match between haplotype groups and regions when examining the entire network. Overall, the data show that a large proportion of genetic variation is shared across Alaska, but this variation is not homogenously distributed across all regions and for all haplotype groups. The second research chapter concentrated on least cisco distributed in freshwater lakes across the Arctic Coastal Plain of Alaska. Little is known about these lake populations, including life history patterns and population genetic structure. We collected least cisco from eight Arctic Coastal Plain lakes, and genotyped six microsatellite loci for each sample. Four of the six loci were not in overall HWE, suggesting that individual lakes may not represent individual subpopulations as initially assumed. Overall, the data from my research suggest that least cisco populations are currently isolated from one another. Isolation also occurred historically, accounting for divergence among major clades. But general recontact events occurred as isolated populations migrated and colonized new habitats, accounting for the heterogeneity found across Alaska. Ultimately, Alaskan least cisco may have functioned as a metapopulation historically, but present populations are too isolated to be considered a metapopulation today.

Appendix

Appendix A1.1. Microsatellite allele frequencies by lake.

Allele	Lake 128	Lake 161	Lake 216	Lake 276	Lake 306	Lake 566	Lake 567
Locus: CLC101							
128	0.179	0.231	0.283	0.262	0.192	0.308	0.000
132	0.643	0.577	0.630	0.738	0.712	0.692	0.000
136	0.179	0.192	0.022	0.000	0.096	0.000	0.000
140	0.000	0.000	0.065	0.000	0.000	0.000	0.000
Locus: CLC7							
286	0.000	0.045	0.000	0.000	0.000	0.000	0.000
290	0.000	0.545	0.357	0.600	0.429	0.000	0.000
294	0.000	0.000	0.036	0.000	0.143	0.000	0.000
298	0.000	0.409	0.607	0.400	0.429	0.000	0.000
Locus: Aut139							
139	0.000	0.024	0.000	0.000	0.000	0.000	0.000
145	0.620	0.810	0.739	0.750	0.962	1.000	0.731
147	0.000	0.024	0.000	0.000	0.000	0.000	0.000
151	0.160	0.048	0.152	0.175	0.038	0.000	0.250
153	0.200	0.048	0.109	0.075	0.000	0.000	0.019
155	0.020	0.048	0.000	0.000	0.000	0.000	0.000
Locus: CLC4							
140	0.018	0.119	0.025	0.000	0.160	0.000	0.000
144	0.518	0.405	0.550	0.625	0.700	0.375	0.000
148	0.411	0.452	0.425	0.350	0.140	0.625	0.000
150	0.000	0.000	0.000	0.025	0.000	0.000	0.000
152	0.018	0.000	0.000	0.000	0.000	0.000	0.000
156	0.000	0.024	0.000	0.000	0.000	0.000	0.000
160	0.018	0.000	0.000	0.000	0.000	0.000	0.000
164	0.018	0.000	0.000	0.000	0.000	0.000	0.000
Locus: CLB129							
295	0.000	0.056	0.000	0.000	0.000	0.000	0.056
299	0.000	0.000	0.000	0.000	0.060	0.000	0.000
303	0.000	0.000	0.000	0.053	0.040	0.000	0.000
307	0.000	0.000	0.021	0.000	0.000	0.000	0.000
311	0.048	0.056	0.000	0.053	0.020	0.000	0.000
315	0.024	0.139	0.063	0.026	0.000	0.000	0.111
319	0.048	0.000	0.000	0.000	0.000	0.033	0.056

Table A1.1 continued

Allele	Lake 128	Lake 161	Lake 216	Lake 276	Lake 306	Lake 566	Lake 567
323	0.071	0.000	0.000	0.000	0.000	0.000	0.000
331	0.024	0.000	0.000	0.000	0.000	0.000	0.000
Allele	Lake 128	Lake 161	Lake 216	Lake 276	Lake 306	Lake 566	Lake 567
Locus: CLB129							
351	0.024	0.028	0.000	0.026	0.000	0.000	0.000
355	0.024	0.000	0.000	0.000	0.000	0.067	0.000
359	0.000	0.000	0.000	0.000	0.000	0.000	0.056
373	0.000	0.028	0.000	0.000	0.000	0.000	0.000
377	0.024	0.000	0.000	0.000	0.000	0.000	0.000
381	0.000	0.028	0.000	0.026	0.020	0.000	0.000
385	0.095	0.028	0.125	0.026	0.000	0.000	0.000
389	0.024	0.028	0.000	0.000	0.000	0.000	0.000
393	0.048	0.139	0.021	0.000	0.040	0.000	0.000
399	0.048	0.000	0.104	0.000	0.000	0.067	0.000
403	0.024	0.000	0.021	0.026	0.040	0.067	0.000
407	0.024	0.083	0.000	0.053	0.140	0.000	0.056
411	0.024	0.028	0.125	0.053	0.120	0.167	0.167
415	0.095	0.056	0.167	0.079	0.020	0.033	0.000
419	0.071	0.056	0.167	0.026	0.060	0.033	0.000
423	0.095	0.028	0.063	0.132	0.020	0.067	0.000
427	0.119	0.194	0.021	0.053	0.300	0.067	0.333
431	0.024	0.000	0.000	0.026	0.000	0.067	0.056
435	0.000	0.000	0.000	0.132	0.060	0.000	0.056
439	0.000	0.028	0.021	0.053	0.040	0.100	0.056
443	0.000	0.000	0.000	0.026	0.000	0.067	0.000
447	0.000	0.000	0.042	0.105	0.000	0.067	0.000
451	0.000	0.000	0.021	0.000	0.020	0.000	0.000
455	0.024	0.000	0.021	0.026	0.000	0.067	0.000
459	0.000	0.000	0.000	0.000	0.000	0.033	0.000